



Facilitating the Development of Game-Based Assessments : An In-depth Exploration of Behavioural Profiling and Soft Skills Recognition via Gamified Situational Judgement Tests

Jérôme Hernandez

► To cite this version:

Jérôme Hernandez. Facilitating the Development of Game-Based Assessments : An In-depth Exploration of Behavioural Profiling and Soft Skills Recognition via Gamified Situational Judgement Tests. Computer Science and Game Theory [cs.GT]. Sorbonne Université, 2023. English. NNT : 2023SORUS589 . tel-04496256

HAL Id: tel-04496256

<https://theses.hal.science/tel-04496256>

Submitted on 8 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

Ecole doctorale Informatique, Télécommunications et Electronique

Laboratoire d'informatique de Paris 6 / MOCAH

Facilitating the Development of Game-Based Assessments: An In-depth Exploration of Behavioural Profiling and Soft Skills Recognition via Gamified Situational Judgement Tests.

Par Jérôme Hernandez

Thèse de doctorat d'informatique

Dirigée par Thibault Carron

Co-encadrée par Mathieu Muratet

Présentée et soutenue publiquement le 11 Décembre 2023

Devant un jury composé de :

- Vanda Luengo, Professeure des universités, Examinatrice (Présidente du jury)
- Yvan Peter, Maître de conférences, Rapporteur
- David Panzoli, Maître de conférences, Rapporteur
- Sébastien Iksal, Professeur des universités, Examineur
- Mohamed Chetouani, Professeur des universités, Examineur
- Thibault Carron, Maître de conférences, Directeur de thèse

Declaration of Authorship

I, Jérôme Hernandez, declare that this thesis titled, ‘Facilitating the Development of Game-Based Assessments: An In-depth Exploration of Behavioural Profiling and Soft Skills Recognition via Gamified Situational Judgement Tests’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Jérôme Hernandez

Date: 11/12/2023

Acknowledgements

I seize this moment to extend my deepest gratitude to everyone who played a pivotal role in my Ph.D. journey. Their priceless support, guidance, and encouragement were crucial in making this research journey successful.

Primarily, I extend my heartfelt thanks to my supervisors, Thibault Carron and Mathieu Muratet, whose mentorship was vital throughout my Ph.D. Their commitment to the project was unwavering, always available when I needed them, providing indispensable expertise and support. They were not just supervisors but mentors who taught me invaluable lessons in becoming a researcher with integrity. They also helped me develop valuable soft skills, enabling me to become a better version of myself than when I started this journey. Their holistic approach to mentorship has not only shaped my career but also my personal development. My sincere appreciation also goes to the jury members: Vanda Luengo, Yvan Peter, David Panzoli, Sébastien Iksal, and Mohamed Chetouani, for their commitment to reviewing my work and offering valuable insights.

I'm profoundly grateful for the camaraderie and support from my colleagues at MOCAH, especially my fellow Ph.D. students, Mélina, Olivier, and Thomas. Not only did we share our challenges and achievements, making this journey unforgettable, but they also consistently took the time to listen to me whenever I needed it. Their willingness to provide thoughtful feedback and insights was invaluable to my Ph.D. journey, helping me to refine my ideas and push my research forward. I wish them continued success and cherish the strong bonds we've formed during this time.

During my tenure at Origamix, although Julie, Marine, Victoire, Julien, Dylan, and finally Tom had brief tenures at the company, our interactions left a lasting impression. Each one, with their keen interest in my research project and unwavering readiness to offer assistance, played a significant role in my journey. I am deeply thankful to them for their support and enthusiasm towards my work. Special gratitude goes to Tom, whose direct involvement in my experiments and studies proved invaluable. His scientific acumen and contributions not only enriched my research but also significantly enhanced my learning and development.

I owe a profound debt of gratitude to my parents, Thérèse and Gregorio, for their unwavering love and support. They went above and beyond, not only by providing the ideal environment for my writing sessions but also by being a constant source of strength and guidance throughout my life. They instilled in us, my brothers, Yohann and Dorian, and I, the importance of striving for excellence in our studies. Their belief in our potential and their relentless encouragement have been the bedrock of my journey,

igniting the passion and determination that led me here. I also want to thank my brothers, Yohann and Dorian, who have been pillars of support, offering encouragement and assistance every step of the way, reinforcing the value of perseverance and resilience that our parents taught us.

At the heart of this journey, two souls stood by me, sharing each day of this endeavor. Tevaite, my partner in life, who has been my unwavering support, and my source of strength. In the labyrinth of research and writing, when the nights stretched long and the deadlines loomed large, it was her belief in me that lit the way. Her presence was a silent promise that I was not alone, that together we could surmount any obstacle. Her sacrifices have not gone unnoticed. The countless times she put my needs above her own, the moments she listened, really listened, to my doubts and fears, and the way she celebrated every small victory as if it were her own - these acts of love and selflessness have been the bedrock of my perseverance.

In the same breath, I also owe a deep sense of gratitude to Ninje, my cherished cat, who was a true companion on this journey. Ninje was there for me through every chapter written and every revision made, a silent yet profoundly comforting presence during those endless nights and days of work. He had this unique way of reminding me to embrace life's simple moments, especially with our daily ritual of walking up and down the stairs, a necessary pause that brought balance to my hectic schedule. Losing Ninje just before my defense was a heartbreak; he didn't get to see the end of our shared adventure. His absence left a void, yet, his spirit and the happiness he brought into my life are forever woven into the fabric of this thesis. To my dear Tevaite and my furry friend Ninje, my heart is filled with gratitude. Your combined spirits and the lessons of love, perseverance, and balance you've taught me will forever be part of who I am.

Thank you, for everything.

Abstract

In recent years, the recognition of non-technical and cross-functional skills has become increasingly common in personnel selection and training. To this end, companies have shown a growing interest in using game-based approaches, coupled with computational data, to predict employee behaviour. Human resources professionals are, therefore, seeking and developing reliable and valid tools to identify the right candidate for the right position while simultaneously enhancing selection strategies and candidate engagement.

This thesis explores the exploration and exploitation of game-based approaches in the field of human resources, particularly in recruitment and selection processes. It first delves into the history of game-based approaches, tracing their origins and explaining their evolution up to today's serious games.

Following that, the thesis presents a meta-frameworks and two frameworks for designing and creating a gamified situational judgement test (GSJT). A systematic progression is highlighted, spanning from grasping client needs to implementing the GSJT. These meta-framework and frameworks undergoes two experimental runs to affirm its effectiveness. Subsequently, a computational psychometric scoring system is introduced to assist GSJT psychometry teams in their scoring endeavors. Several scoring methods are scrutinized, and a hybrid scoring approach is introduced. A third experiment is executed to validate the efficacy of this approach.

The thesis then shifts its focus to the significance of having lifelike non-playable characters for recognizing non-technical skills. Deep learning models are suggested to facilitate the creation of these characters. A final experiment is conducted to gauge the effectiveness of these models.

In conclusion, this thesis provides preliminary contributions to both theory and practical applications concerning the creation and use of game-based methodologies in the human resources sector for behavioural profile recognition.

Contents

Declaration of Authorship	i
Acknowledgements	ii
Abstract	iv
List of Figures	x
List of Tables	xii
Abbreviations	xiii
1 Introduction	1
1.1 Research Context	1
1.1.1 What are Soft Skills?	1
1.1.2 Importance of Soft Skills and Behavioural Profile Recognition	3
1.1.3 Soft Skills Measurements	5
1.1.4 Psychometrics Models	7
1.2 Industrial Agreements for Training for Embedding Research Issues in Real-World Contexts.	9
1.3 Financial Implications of Creating Game-Based Approaches	10
1.4 Contextual Particularities and Selected Direction: Defining our Research Subject	12
1.5 Research Questions, Hypotheses, and Objectives	14
1.5.1 Hypotheses related to Objective A: A Game-Based Approach Model which Foster Collaboration	14
1.5.2 Hypotheses related to Objective B: Streamline the gamifi- cation	15
1.6 Research Methods	16
2 Game-based Approaches History	20
2.1 Game-based Approaches' Origins	21
2.2 Introduction to Serious Games	24

2.3	Serious Games are not only Digital	25
2.4	Simulation Games are Serious Games	26
2.5	Edutainment Games, the Game-Based Learning Precursors	27
2.6	Serious Gaming and Serious Game	29
2.7	Digital Game-Based Learning	32
2.8	Play and Game Inspired Design: Overview	33
2.9	Gamification and Toyification	35
2.10	Conclusion	40
3	Related Works	43
3.1	Human Resources and Game-based Approaches for Recruitment and Selection Processes	44
3.1.1	Game-based Assessment to Enhanced Candidate Experience	46
3.1.2	Game-based Assessment to Enhanced Candidate Pool	46
3.1.3	Game-Based Assessment to Applicant Reactions to Selec- tion Process	49
3.1.4	Objective measurement in Game-Based Assessment	51
3.1.5	Game-Based Assessment to Address Bias	52
3.1.6	Conclusion	55
3.2	Game-based Soft Skill Assessment	55
3.2.1	Simulations for Recruitment and Selection	56
3.2.2	Gamified Situational Judgement Tests	59
3.2.3	Interactive Narrative Game and Gamified Situational Judge- ment Test	62
3.3	Toward the Design and Conception of GSJT	63
3.3.1	Stakeholders involved in Situational Judgement Test	64
3.3.2	Design and Conception of Situational Judgement Tests	65
3.3.3	Design and Conception of Interactive Narrative Games	66
3.3.3.1	Drama Manager in Interactive Narrative Games	68
3.3.3.2	Non-Playable Characters in Interactive Narrative Games	69
3.3.3.3	Player Agency, Assessment System and Interactive Storytelling	70
3.3.4	Stakeholders involved in Interactive Narrative Games	72
3.4	Conclusion	74
4	Proposing a Meta-Framework and a Preliminary Framework to Generate a Gamified Situational Judgement Test	78
4.1	Meta-framework for designing and creating Gamified Situational Judgement Test	79
4.2	Toward a first framework: Identification and elaboration of the stages involved in the development of a GSJT	84
4.2.1	Stage 1: Defining the client request	85

4.2.2	Stage 2: Generating various traditional SJT situations	87
4.2.3	Stage 3: Screenwriting for INGs	89
4.2.4	Stage 4: Crafting the gamification based on ING	90
4.2.5	Stage 5: Implementing the GSJT	91
4.2.6	Conclusion	92
4.3	Experimentation 1: A first GSJT	92
4.3.1	Stage 1: Experiment Objectives	93
4.3.1.1	Stage 1: Experimental Methodology	93
4.3.1.2	Stage1: Experimental GSJT design and development	93
4.3.1.3	Stage 1: Results	95
4.3.2	Stage 2: Experiment Objectives	97
4.3.2.1	Stage 2: Participants	97
4.3.2.2	Procedure	98
4.3.2.3	Stage 2: Results and discussion	98
5	Proposing a Second Framework based on a Modular Model, Conjugated with a Drama Manager	104
5.1	Semi-Emergent Modular Gamified Situational Judgement Test Approach	106
5.1.1	Narrative modules	107
5.1.2	Interfaces	108
5.2	Global Semi-Emergent Modular GSJT Design Approach	109
5.2.1	GSJT stakeholders' and Roles Identification	109
5.2.2	Modular Production Workflow	112
5.3	A Drama Manager to Dynamically Generate a GSJT	114
5.3.1	Global Game State	115
5.3.2	Narrative Module Library	116
5.3.3	Narrative Module Selector	116
5.3.4	Narrative Module Adaptation Mechanism and Narrative Feedback System	120
5.3.5	Conclusion	121
5.4	Experiment 2: A second GSJT, Experimental Design and Objectives	121
5.4.1	The Experimental Gamified Situational Judgement Test and SJT	122
5.4.2	Sample and Procedures	123
5.4.3	Results and Discussion	124
5.4.4	Tests of the hypothesis	125
5.4.5	Conclusion	127
6	Proposing a Computational Psychometric Scoring System to Help GSJT's Psychometrics Teams	130
6.1	Scoring a Situational Judgment Tests	131

6.1.1	Scoring Methods: Modalities, Advantages and Limitations	132
6.1.2	Conclusion	136
6.2	A Computational Psychometric Scoring Approach	136
6.2.1	Scoring Matrices, Behavioural Profile and Scoring System . .	137
6.2.2	A Hybrid scoring method approach	140
6.3	Experimentation 3: Computational Psychometric Scoring System Implementation, Experimental Design and Objectives	144
6.3.1	Experimental Gamified Situational Judgment Test	144
6.3.2	Sample and Procedures	145
6.3.3	Results and Discussion	146
6.4	Conclusion	150
7	Machine and deep learning models to help the creation of realistic non-playable characters	152
7.1	Why it's important to have realistic characters for soft skill recognition	152
7.2	Proposing dialogues personality classifiers	153
7.3	Creation of a binary classifier for Non-Playable Characters Dialogues	154
7.3.1	Stage 1: Document collection and labelization	155
7.3.2	Stage 2: Document Preprocessing	156
7.3.3	Stage 3: Train and Generate the Models	158
7.4	Experimentation 4: Development and implementation of NPC's binary personality classifiers, Experimental Design and Objectives	159
7.4.1	Psychometry, Tools and Dataset	159
7.4.1.1	MBTI and Big Five Typologies	160
7.4.1.2	Pre-study on the Correlation between the Big Five and MBTI typologies	161
7.4.1.3	Experiment Tools	162
7.4.1.4	Game Pool & Dataset	163
7.4.2	Experimental Method	164
7.4.2.1	Dataset Analysis	164
7.4.2.2	Data Preprocessing	166
7.4.2.3	Model Training	167
7.4.3	Supervised Machine Learning Algorithms	167
7.4.4	Logistic Regression Algorithm	168
7.4.5	Support Vector Machine Algorithm	168
7.4.6	Naïve Bayes Algorithm	169
7.4.7	Decision Tree Algorithm	170

7.4.8	Random Forest Algorithm	171
7.4.9	Deep Learning Algorithm Using Convolutional Neural Network	172
7.4.10	Results and Discussion	174
	7.4.10.1 Supervised Learning Methods Comparison	174
	7.4.10.2 Limits of the study	175
7.4.11	Conclusion	176
8	General Conclusion	179
8.1	Our Contribution	179
8.2	Perspectives and Future Research	181
	Bibliography	186
	Experimentation 1: Survey Questions	209

List of Figures

1.1	Estimated Total Hours Worked in Europe and the United States, Comparison Between 2016 and 2030, in Billions ¹	4
1.2	Our methodological approach	17
2.1	Conceptual map of Game-Based Approaches by Deterding et al. (2016)	21
2.2	Game of Go	22
2.3	A Kriegsspiel Match in Progress (Photo by Kurzon - Wikipedia)	23
2.4	Oregon Trail Game	28
2.5	Math Blaster, Where in the World is Carmen Sandiego? and Reader Rabbits Games	30
2.6	Zombies, Run! by Six to Start	37
2.7	Fitbit Gamification Example by Fitbit company	37
2.8	Re-Mission by HopeLab Foundation	38
2.9	The Lego Serious Play by The Lego Group	39
2.10	Minecraft: Education Edition by Microsoft	40
3.1	Reveal Serious Game by L'Oréal	47
3.2	America's Army: Proving Grounds by United States Army	48
3.3	My Marriott Motel by Marriott Hotels	48
3.4	Model of Applicants' Reactions to Employment Selection Process and Outcome (from Gilliland, 1993, p. 700)	50
3.5	Suggested impacts of gaming elements on faking in game-based assessment. (from Landers and Sanchez, 2022, p. 9)	51
3.6	Circuit Runner game. The left division portrays the player's navigation and interactivity within the labyrinth, whereas the right is subdivided into two segments. The upper subsection (A) showcases the conversational history between the player and the virtual agent. The lower subsection (B) presents the player's possible responses and prompts a numeric selection of these choices. [1]	53
3.7	Example of a Gamified Situational Judgement Test for soft skill recognition by Owiwi ²	59
3.8	Example of Interactive Narrative Games: Life is Strange, The Walking Dead, Danganronpa)	62
3.9	Situational Judgement Test Design and Production Process (From Patterson and Driver, 2018, p. 82)	65
3.10	Narrative Structure Categories	67

3.11	Facade game by Mateas and Stern	71
4.1	Gamified Situational Judgement Test Meta-Framework	81
4.2	Gamify situational judgement test production timeline	83
4.3	Gamified Situational Judgement Test Design Process in 5 Steps	85
4.4	Narrative Module Internal Structure Example	95
4.5	Screenshot of the Result GSJT	96
4.6	Clustering evolution of the forty-eight players	100
4.7	Elbow method and K-means clustering on the sixty-five players after the fifth game situation	101
4.8	Players distribution by game score and soft skills	101
5.1	Traditional and Modular Gamify Situational Judgement Test	106
5.2	Narrative Modular combination	109
5.3	Gamified Situational Judgement Test Design Process in 6 Steps	110
5.4	Modular Production Workflow	113
5.5	Dynamic Narrative Modular System	115
5.6	Four Examples of the Narrative Module Selection	117
5.7	Example of an Assessment Module which assess "Inquisitiveness" soft skill in the Experimental GSJT Note: Character sprite and back- ground are not from the actual game and are adapted from free-to-use assets available on Freepik and created by 'upklyak' sur Freepik for illustration purposes.	123
6.1	Example of Three Behavioural Profiles	138
6.2	An example of a Narrative Module Decision Graph	139
6.3	Psychometric Team Approach to Generate a Behavioural Profile	141
6.4	Computational Psychometrics Approach to Generate a Behavioural Profile	142
6.5	Computational Psychometrics Algorithm Process	143
6.6	Euclidian analysis between three pairs of behavioural profiles	147
6.7	Euclidian Analysis Box-plot Visualization for the Both Scoring System	148
7.1	Personality Classifier Generation Process	155
7.2	MBTI Dataset Distribution Analysis	165
7.3	Classification process using CNN approach with spaCy Library	173

List of Tables

4.1	Cooperation Item's Answer Scoring Grid	89
5.1	Intercorrelations for all variables	125
5.2	Means, standard deviations, and internal consistency reliability . . .	126
6.1	Example of a Rating Scale	138
6.2	Scoring Grid	145
7.1	Correlation between the Big Five traits and the MBTI traits	161
7.2	Interpretation Cohen's Kappa value	162
7.3	MBTI's Traits Distribution	166
7.4	Big Five's Traits Distribution	166
7.5	Logistic Regression Classifiers Evaluation	168
7.6	Support Vector Machine Classifiers Evaluation	169
7.7	Naive Bayes Classifiers Evaluation	170
7.8	Decision Tree Classifiers Evaluation	171
7.9	Random Forest Classifiers Evaluation	172
7.10	Convolutional Neural Network Classifiers Evaluation	173
7.11	Methods Comparaison for the MBTI Typology	174
7.12	Methods Comparaison for the Big Five Typology	174
7.13	Top classifiers by traits	175

Abbreviations

HR	H uman R esources
SSk	S oft S kill
GBA	G ame- B ased A pproaches
SJT	S ituational J udgement T est
GSJT	G amified S ituational J udgement T est
ING	I nteractive N arrative G ames
NPC	N on- P layable C haracter
DM	D rama M anager
RBP	R eferential B ehavioural P rofile
GGs	G lobal G ame S tate
NFS	N arrative F eedback S ystem
FFT	F ive- F actor M odel
RV	R esponse V ectors
CP	C omputational P sychometrics
L2	Euclidean distance
NLP	N atural L anguage P rocessing
LR	L ogistic R egression
SVM	S upport V ector M achine
DT	D ecision T ree
NB	N aive B ayes
RF	R andom F orest
CNN	C onvolutional N eural N etwork
MBTI	M yers- B riggs T ype I ndicator

OCEAN	O penness, C onscientiousness, E xtraversion, A greeableness, and N euroticism
SME	S ubject M atter E xperts
SDT	S elf- D etermination T heory
MMORPG	M assively M ultiplayer O nline R ole- P laying G ame
FPS	F irst- P erson S hooter
UCD	U ser- C entered D esign

Chapter 1

Introduction

1.1 Research Context

In recent years, the fields of **gamification and serious games** have experienced substantial growth, thanks to the integration of cutting-edge technology and innovative design principles [2, 3]. This surge in interest has led to the application of **game-based approaches** across various domains, including education, health, and business [4–6]. The purpose of this PhD thesis is to explore the potential of gamification and serious games in the realm of **human resources**, with a specific focus on **soft skills** and **behavioural recognition** in **recruitment and selection process**.

1.1.1 What are Soft Skills?

The exact definition of '**soft skills**' has long been a subject of debate within the academic community, largely due to its layered and **multifaceted nature** and its roots embedded in diverse disciplines [7]. The term often serves as a broad encompassing category, covering a wide range of non-technical abilities, competencies, and attributes that are seen as vital for achieving success in both personal and professional life [8].

The inherent multidimensionality of soft skills contributes to the difficulty in pinning down a universally accepted definition. Depending on the perspective from which they are viewed, soft skills can be conceptualised as **personality traits**,

interpersonal and social skills, or cognitive abilities, among others [9].

Each of these perspectives provides a unique interpretation of soft skills, but none is comprehensive enough to capture the full breadth and depth of the concept.

Firstly, from the perspective of *personality traits*, soft skills are considered as **innate attributes** that influence an individual's behaviour, interactions, and reactions to different situations. For example, traits like conscientiousness, emotional stability, and openness to experience have been found to have significant impacts on job performance and career success [10]. However, this perspective can be criticised for its relatively static view of personality traits as soft skills, as it may not adequately account for the potential for growth and development in these skills over time [9].

Secondly, when viewed as *interpersonal skills*, soft skills are seen as primarily **social abilities**. This perspective emphasises the relational and communicative aspects of soft skills, such as effective communication, teamwork, and conflict resolution [11]. While these skills are undeniably crucial, this interpretation may not adequately account for non-social components of soft skills, such as resilience and self-regulation [12].

Finally, the perspective that views soft skills as *cognitive abilities* emphasises **mental processes**, including critical thinking, problem-solving, and decision-making [13]. While these cognitive skills are an essential part of the broader soft skills framework, an overemphasis on this perspective might undervalue the importance of emotional and social competencies, which are equally crucial in many personal and professional situations [9].

The variety of interpretations of soft skills reflects their complexity and underscores the challenges faced by researchers, educators, and practitioners in defining and operationalising this concept. However, this definitional ambiguity also attests to the richness and breadth of the concept, highlighting the diverse range of human capabilities it embodies, extending beyond the boundaries of personality traits, interpersonal interactions, and cognitive processes.

In this thesis, for the sake of clarity and consistency, we use the term '**soft skills**' as an overarching term to cover all related interpretations and aspects. This decision recognises and respects the multifaceted nature of soft skills and their importance in recruitment and selection processes. It is important to note, however, that the usage of the term '*soft skills*' in this context does not negate or

simplify the complexities associated with defining this term. Instead, it serves as a pragmatic approach that facilitates a cohesive exploration and understanding of this complex and multifaceted concept in our context of recruitment and selection.

1.1.2 Importance of Soft Skills and Behavioural Profile Recognition

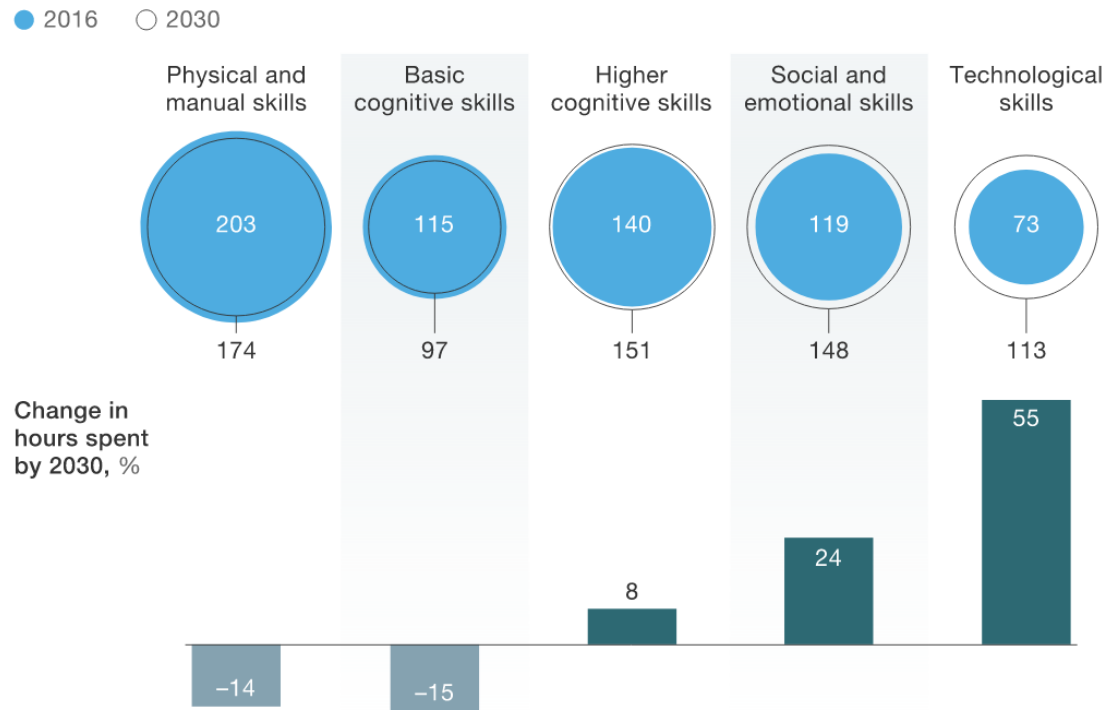
In today's competitive job market, soft skills have emerged as vital factors in determining an individual's success and employability [9, 14]. Soft skills, unlike hard skills, which pertain to technical and job-specific abilities, may refer to personality and behavioural traits and to non-technical, interpersonal, and social skills that enhance an individual's performance across various industries and roles [7, 15]. These skills, which are less tangible and more difficult to quantify, have gained increasing prominence in the modern workplace.

Indeed, communication, adaptability, and emotional intelligence are some examples of soft skills that play a crucial role in fostering effective workplace relationships and enhancing overall performance [8, 16]. Other essential soft skills include problem-solving, teamwork and collaboration, time management, and leadership, which contribute to an individual's ability to navigate complex work environments and drive organisational success [14].

The shift towards knowledge-based economies and the impact of automation have necessitated a greater emphasis on soft skills in the workforce. Manyika and Sneider [17] suggest in their study that as routine tasks become automated, employees will need in the near future to demonstrate advanced cognitive, social, emotional and interpersonal skills to remain competitive in the job market (see Fig 1.1).

Adaptable and versatile employees are highly sought after in rapidly changing work environments, as they can quickly adjust to new technologies, processes, and organisational structures. Furthermore, emotional intelligence and empathy have gained importance in diverse teams, as they facilitate cross-cultural communication and foster a positive work environment [16].

¹Source: McKinsey Global Institute Workforce Skills Model; McKinsey Global Institute analysis



Source: McKinsey Global Institute Workforce Skills Model; McKinsey Global Institute analysis

FIGURE 1.1: Estimated Total Hours Worked in Europe and the United States, Comparison Between 2016 and 2030, in Billions ¹

Employers increasingly recognise the importance of soft skills and integrate their assessment into recruitment and selection processes [9, 18]. On one hand, recruitment is the process of **attracting, engaging, and encouraging qualified individuals** to apply for job openings within an organisation [19]. The primary objective of recruitment is to create a pool of potential candidates from which the organisation can select the most suitable individuals for the job [20]. Recruitment methods often involve a combination of job advertisements, talent sourcing, campus recruitment, and job fairs.

Selection, on the other hand, is the process of **evaluating candidates** who have applied for a job to determine which individuals best fit the requirements of the position and the organisation's culture [21]. The main objective of selection is to predict future job performance and minimise the risk of a poor hire [22]. Selection methods typically involve a combination of assessments, such as interviews, cognitive ability tests, and personality assessments [23].

Subjectivity and bias in evaluation, as well as the difficulty in quantifying soft

skills, present challenges in their assessment during recruitment and selection processes [9]. These challenges necessitate the development of reliable and valid methods for assessing candidates' soft skills.

Incorporating soft skills in recruitment and selection processes can lead to improved employee performance, **enhanced team dynamics, and increased job satisfaction** [9, 14]. Moreover, the focus on soft skills can contribute to a reduction in employee turnover and associated costs, as well as foster organisational adaptability and resilience in the face of rapid changes and challenges [24].

In conclusion, the growing importance of soft skills in the job market underscores their significance in recruitment and selection processes. By integrating soft skills assessment into hiring strategies, organisations can better identify individuals who possess the right mix of technical and interpersonal abilities to succeed in their roles.

1.1.3 Soft Skills Measurements

Soft skills, despite their noted importance, pose significant measurement challenges for researchers and practitioners, largely attributed to their subjective, multifaceted, and context-dependent nature [7]. Soft skills are difficult to quantify, and their manifestation can greatly vary depending on the situational context, personal dispositions, and cultural influences [25]. Therefore, designing a universally applicable metric for their measurement is no small task. Various methods have been proposed and utilised for assessing soft skills, each offering distinctive advantages but also possessing inherent limitations, necessitating the use of a combination of methods for a more holistic evaluation [26]. We present in the following a non-exhaustive list of approaches used to measure soft skills.

- **Self-report questionnaires** are a prevalent method for assessing soft skills, primarily due to their **cost-effectiveness and streamlined data gathering**. These questionnaires typically involve a series of statements to which individuals respond, reflecting their self-perceived proficiency in various soft skills. However, self-report measures have been criticised for their susceptibility to bias. Respondents may tend to portray themselves favourably,

providing socially desirable rather than truthful responses [27]. Additionally, individuals may lack accurate self-perception, leading to inconsistencies between self-reported and actual skill levels [28].

- **Observer ratings**, which include assessments by peers, teachers, or supervisors, offer an alternative method for assessing soft skills. Unlike self-reports, observer ratings provide an **external perspective** on an individual's soft skills, potentially offering more objective insights [29]. However, these assessments are not without their limitations. The accuracy of observer ratings can be influenced by various factors, including the rater's personal bias, the quality of their relationship with the person being rated, and their exposure to the individual's behaviour across different contexts.
- **Performance tasks**, frequently used in the assessment of cognitive soft skills such as critical thinking or problem-solving, involves participants performing specific tasks while their performance is evaluated [9]. This method offers a more **direct measure** of soft skills, as it allows for the observation of these skills in action. Despite its benefits, the use of performance tasks can be resource-intensive, requiring considerable time and expertise to develop, administer, and evaluate. Furthermore, performance tasks often focus on specific skills in isolation, potentially overlooking the integrated and context-dependent nature of soft skills.
- **Behavioural simulations**, such as role-plays or group exercises, present another approach to the assessment of soft skills. These simulations provide a **dynamic and interactive context** for individuals to demonstrate their soft skills in response to realistic scenarios [30]. However, the artificiality of these scenarios and the awareness of being evaluated may influence participants' behaviour, potentially compromising the ecological validity of the assessment [31].

In conclusion, while the measurement of soft skills is fraught with significant challenges due to their complex and multifaceted nature, adopting a multi-method approach that leverages a combination of assessment techniques can offer a more comprehensive and robust evaluation of these competencies [26]. Therefore, organisations may use psychometrics models to help them in their behavioural recognition endeavour.

1.1.4 Psychometrics Models

Among the various methods used to assess potential employees, **personality models** have gained prominence due to their ability to predict job performance and success across a range of occupations and industries [10, 32]. This thesis relies on three of the most widely used personality models in companies for recruitment and selection, namely the Big Five Personality Traits, also known as the *Five-Factor Model (FFM)*, the *Myers-Briggs Type Indicator (MBTI)*, and the *HEXACO model* [33].

To begin with, the Big Five Personality Traits or the FFM is a widely accepted model that identifies **five key personality traits**: *Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN)* [34]. The FFM has garnered substantial empirical support and is regarded as a reliable tool for predicting job performance and success [10, 32]. Moreover, this model's applicability extends across various cultures and languages, making it a versatile instrument for assessing individuals in diverse contexts [35].

However, despite its numerous strengths, the FFM has been criticised for certain limitations. For instance, Block [36] argued that the model may not capture the full complexity of human personality, while McAdams [37] claimed that its reductionist nature confines individuals to just five traits, which may not accurately represent their unique personalities.

In contrast to the FFM, the MBTI categorises individuals into **16 distinct personality types** based on four dichotomous dimensions: *Extraversion-Introversion, Sensing-Intuition, Thinking-Feeling, and Judging-Perceiving* [38]. This model is popular in organisations for purposes such as team building, leadership development, and career counselling [39, 40]. One of its major advantages is its ease of understanding and application in organisational settings, making it a practical tool for promoting self-awareness and team development [41, 42].

Nonetheless, the MBTI has received criticism for its lack of empirical support and reliability when compared to the FFM [40, 42]. Furthermore, the dichotomous nature of the dimensions in the MBTI has been criticised for oversimplifying human personality, leading to concerns about the accuracy of the model [43].

Finally, the HEXACO model is another prominent personality model that has emerged in recent years as an alternative to the FFM and MBTI. The HEXACO

model identifies **six key personality dimensions**: *Honesty-Humility (H)*, *Emotionality (E)*, *Extraversion (X)*, *Agreeableness (A)*, *Conscientiousness (C)*, and *Openness to Experience (O)* [44]. This model has gained traction in the academic community due to its ability to predict work-related behaviours and outcomes, as well as its cross-cultural applicability.

One of the primary advantages of the HEXACO model is its inclusion of the *Honesty-Humility dimension*, which distinguishes it from the FFM. This additional dimension allows for a more comprehensive understanding of individual personalities by capturing aspects related to integrity, sincerity, fairness, and modesty. Furthermore, Ashton et al. [44] argue that the HEXACO model has demonstrated strong psychometric properties, including reliability and validity, making it a robust tool for predicting job performance, team dynamics, and leadership potential. The researchers also suggest another advantage of the HEXACO model is its cross-cultural applicability, which has been demonstrated through numerous studies conducted across different countries and languages. This suggests that the HEXACO model may be particularly useful for organisations with a global presence or those that are culturally diverse.

Despite its strengths, the HEXACO model also has some limitations. One of the concerns is that it may be more complex than the FFM due to its inclusion of an additional dimension, which could make it more challenging for practitioners to apply and interpret [44]. Additionally, although the HEXACO model has gained increasing recognition and support in the academic literature, it is still relatively less established compared to the FFM and MBTI.

In conclusion, the FFM, MBTI, and HEXACO models each have their own merits and drawbacks as personality models used for recruitment and selection in companies. The FFM boasts strong empirical support, reliability, and applicability across cultures but has been criticised for its reductionist nature and potential failure to capture the complete range of human personality traits [36, 37]. On the other hand, the MBTI is easy to understand and apply but lacks empirical support and reliability compared to the FFM, raising questions about its overall effectiveness in assessing individual personalities [40]. Finally, the HEXACO model, with its additional Honesty-Humility dimension, offers a more comprehensive understanding of personality, demonstrating strong psychometric properties and cross-cultural applicability, but its complexity and relative novelty may limit its widespread adoption.

Although each model possesses its own unique set of strengths and limitations, their continued use in companies can be attributed to factors such as empirical support, practicality, and ease of implementation. In addition, despite the ongoing debate within the academic community regarding the validity and comprehensiveness of these models, organisations continue to rely on them for assessing individual characteristics and aptitude for various roles. Therefore, we base our research on the three following personality typologies in order to follow the thesis objectives while remaining mindful of the potential limitations inherent in the reductionist nature of personality typologies.

Finally, organisations not only tend to propose reliable and coherent selection methods but also tend to propose a **great experience** for the candidate for recruitment purposes. Therefore, a balance between recruitment and selection is required by organisations.

Building on this foundation, our contribution intends to situate itself within the ongoing exploration of these personality models. Our aim is to be tool-independent, striving for a generic methodological approach that can be applied, whether it's through a model created by an individual based on a questionnaire or any other existing tool.

By default, we utilise what already exists in the realm of personality assessment. However, the heart of our contribution lies in the capacity to adapt to any tool. Our work is designed to be *flexible and universally applicable*, adding to the versatility of personality assessment. This adaptability may enhance the utility of personality assessment in recruitment and selection processes, facilitating the match between individuals and organisations in a more efficient and effective manner.

1.2 Industrial Agreements for Training for Embedding Research Issues in Real-World Contexts.

This thesis work commenced under the CIFRE² agreement at Origamix, a startup with limited resources. The research was informed by observed practices and field

²Industrial Training and Research Agreement: <https://www.anrt.asso.fr/fr/le-dispositif-cifre-7844>

issues identified through various client projects within the company. However, due to company reasons, the agreement was terminated at the end of the second year of the thesis. Consequently, the final year of the thesis took place within the LIP6 in the MOCAH team.

Origamix specializes in gamification for human resources. Its core business involved designing and producing physical serious games ³ for clients from diverse sectors (industry, banking and insurance, clothing, services) and various sizes (small to large-scale companies).

The evolutions in the human resources market and need have led the company to take an interest in new innovative approaches, particularly in the use of digital games for recruitment and selection purposes. Indeed, Origamix receives several inquiries to create various serious games with a short game duration to evaluate a large volume of diverse player behavioural profiles. Hence, the research conducted in this thesis was inspired by these innovative requirements and highlights the company's ambition to discover alternative applications of serious games for recruitment and selection purposes. The objective is to capitalize on the numerous advantages of digital serious games and technologies, transcending traditional assessments or digital tests by developing game-based assessments for recruitment and selection processes.

1.3 Financial Implications of Creating Game-Based Approaches

One major concern in creating game-based assessments is the considerable financial investment required to develop and implement these game-based tools. The development of a game-based approach necessitates significant financial investments in design, programming, and testing to ensure the effectiveness, reliability, and validity of the assessment. For example, Marfisi-Schottman [45] highlighted that the production cost for *America's Army*⁴ game [46], one of the first games

³A serious game is a type of interactive digital or physical game designed to achieve specific learning, training, or problem-solving objectives while engaging and entertaining its players. (see extensive definition at 2.2)

⁴America's Army is a series of first-person shooter (FPS) video games developed by the United States Army as a recruitment, communication, and training tool.

used for recruitment purposes, was more than 30 million American dollars. The researcher also estimates that serious games may cost between 10 to 300,000 dollars, based on the study of Aldrich and Millet [47, 48].

Firstly, creating a game-based approach requires a **multidisciplinary team of professionals**, including game designers, software developers, and subject matter experts, to design engaging game scenarios and align them with the assessment criteria [49]. Moreover, the development process involves iterative testing and refining the game-based approach to guarantee accurate and reliable results [50]. The initial development costs for a game-based approach can be substantial, especially for smaller organisations or those with limited budgets. As a result, the associated development costs may hinder its widespread adoption and present a significant financial challenge for organisations considering this approach.

In addition, customizing a game-based approach to meet the specific requirements of an organisation and its recruitment and selection process adds to the overall cost of creating these assessment tools. Indeed, customization may involve tailoring the game's content, difficulty levels, and assessment criteria to align with the organisation's industry, job roles, and desired candidate competencies [51]. Consequently, the additional costs associated with customization contribute to the overall financial burden of adopting a game-based approach for recruitment and selection purposes.

Finally, maintaining a digital game-based approach tool requires consistent investments in technical support to address software issues, ensure system compatibility with evolving technology, provide user assistance, and evolve the game with the market evolution. Additionally, game-based approach tools must be periodically updated to remain relevant and engaging, necessitating further investments in design, programming, and content development. The continuous maintenance and update costs for a game-based approach may strain an organisation's resources, potentially making it a less viable investment option compared to traditional assessment methods. Thus, the long-term costs of maintaining and updating game-based approach tools present an additional financial challenge for organisations considering this approach for recruitment and selection processes.

In conclusion, the creation of game-based assessments for recruitment and selection processes involves **substantial financial investments in development, customization, and maintenance**. The high costs associated with each of these

stages may present a significant barrier to the adoption of a game-based approach, particularly for smaller organisations or those with limited budgets. Moreover, additional challenges emerge from constraints that are delineated in the subsequent section. In the course of this thesis, we aim to propose **strategies and solutions that could potentially alleviate these barriers and challenges**, thereby fostering more widespread use of game-based assessments in various organisational contexts.

1.4 Contextual Particularities and Selected Direction: Defining our Research Subject

In our context, addressing the recurring characteristics and needs of companies and organisations seeking game-based assessments is essential. In the following, we aim to analyse and address several key challenges faced by these organisations in designing and implementing assessments:

- **Customization and adaptability:** Organisations and companies often require assessments tailored to their **specific professional context, industry, or organisational culture**. As a result, assessments should be designed with flexibility in mind, allowing for customisation to suit various contexts and address unique challenges faced by different stakeholders.
- **Engagement and motivation:** In order to maintain stakeholder interest and ensure effective soft skill and behavioural assessment, it is crucial to create game-based assessments that are both **engaging and motivating**. This can be achieved by incorporating game elements, such as narratives, challenges, and rewards, that resonate with the target audience and stimulate their curiosity and desire to participate in the recruitment and selection process.
- **Real-world relevance:** Organisations and companies generally look for assessments that align with their specific professional contexts when recruiting and selecting candidates. These assessments should enable candidates to demonstrate how their skills and knowledge can be applied directly to their potential job responsibilities. Consequently, it is crucial to create assessments for recruitment and selection purposes that **emulate real-world**

situations and challenges, promoting the evaluation of practical soft skills in candidates.

- **Scalability and accessibility:** As organisations and companies may operate on different scales, both nationally and internationally, it is crucial to design assessments that **accommodate candidates from various geographical locations and technical backgrounds**. This can be achieved by ensuring that the assessments are accessible through multiple platforms and devices, as well as providing support for different languages and cultural contexts. Furthermore, it is essential to create inclusive assessments tailored for recruitment and selection that consider the diverse socio-cultural backgrounds of potential candidates. This promotes equal opportunity and accessibility for all applicants.
- **Collaboration and social interaction:** Professional contexts often involve collaborative work and interpersonal communication. Assessment means should also assess soft skills such as teamwork, knowledge sharing, and the co-creation of solutions by incorporating collaborative tasks and opportunities for social interaction. However, since assessment means need to be accessible to candidates individually and at any time, the design should include situations that **emulate teamwork scenarios**.
- **Data-driven insights and feedback:** To optimise the assessment experience and drive continuous improvement, organisations and companies require assessments that provide **actionable insights and feedback**. Assessments should be designed to collect and analyse relevant data, allowing stakeholders to identify trends, track progress, and make informed decisions about future recruitment and selection initiatives.
- **Time-efficient and concise:** It is essential to design assessment means to be **time-efficient and concise**. Candidates and hiring managers often have limited time available during the recruitment process, so assessments should be designed to quickly and effectively evaluate the necessary skills and competencies without causing unnecessary delays.
- **Cost-effective and budget-friendly:** As explained in the previous paragraph, human resources departments often face financial constraints and limited budgets for recruitment and selection processes. As a result, it is crucial to design assessments that are **cost-effective and budget-friendly**,

as well as offer scalable pricing models that accommodate various organisational sizes and resource allocations.

1.5 Research Questions, Hypotheses, and Objectives

Drawing from the observations and requirements discussed in earlier sections, and in response to Origamix's request, our research concentrates on one hand on proposing a **game-based approach** followed by production frameworks and a model to create a serious game for soft skill and behavioural recognition. On the other hand, we focus our research on **assisting the design and production of a game-based approach** based on the proposed model. Consequently, we have identified two primary research objectives that we intend to explore and address throughout this thesis:

- **Objective A:** Propose a digital game-based approach model to foster an immersive gaming and assessment environment. With the aim to enhance understanding of applicants' behaviours, personalities and organisational fit, all while promoting heightened engagement and sustained motivation among participants.
- **Objective B:** Propose strategies and innovative solutions to facilitate and assist a game-based assessment creation and gamification process that attains similar efficacy and behavioural identification results as conventional recruitment and selection techniques.

In the next paragraph, we introduce the hypotheses that we set for this thesis, aiming to identify effective solutions to address the dual research objectives we have outlined.

1.5.1 Hypotheses related to Objective A: A Game-Based Approach Model which Foster Collaboration

The primary research question focuses on identifying and proposing a game-based approach model for behavioural recognition. On the other hand, we also focus on

enhancing collaboration among the various stakeholders involved in developing a game-based approach. Our goal is to provide these stakeholders with a clear and well-defined understanding of their tasks and the sequence in which they must be executed in order to streamline the collaborative process.

In order to address this question, we have formulated several hypotheses, which will be discussed in detail in Chapters 4 and 5:

- Is our game-based approach model proposal enabling the generation of an immersive serious game to recognise soft skills for the recruitment and selection process?
- Can the development and application of a meta-framework facilitate the establishment of specialized frameworks that provide a clear delineation of roles and responsibilities of all parties involved in the design and production process? This approach may enhance clarity and effectiveness in the game-based behavioural recognition process.

1.5.2 Hypotheses related to Objective B: Streamline the gamification

The second research question addresses the assistance in enhancing the gamification process for traditional recruitment and selection methods.

To address this question, we have formulated several assumptions that we will discuss in Chapters 5, 6 and 7:

- A dynamic narrative interactive approach may propose an efficient way to enable a unique experience and foster replayability of a game-based approach for recruitment and selection purposes.
- The application of computational psychometrics may assist psychometric experts in developing a relevant in-game soft skill measurement system.
- By employing machine learning and deep learning techniques, it is possible to classify non-playable character (NPC) personalities in order to enrich and heighten the significance of in-game interpersonal interactions.

1.6 Research Methods

Our investigation is grounded in a thorough review of the history of research and application domains related to using game-based approaches for serious purposes, as presented in Chapter 2. This exploratory and foundational study paves the way for establishing distinct research phases. These phases are designed to facilitate the development of proposals that address the overarching goal: *to support the design and creation of a game-based approach tool for recognising behavioural and soft skills within the context of selection and recruitment processes.*

Following the game-based approaches exploratory study, our research was divided into three separate work phases, which guided the development of proposals aimed at addressing the identified problem (see Figure 1.2).

- The initial phase of our project is devoted to comprehensive theoretical investigations, which serve as both an exploratory and foundational structure. This study enables us to gain a deeper comprehension of the multifaceted domain in which our research is situated, ultimately leading to a structured representation. Subsequently, our focus narrows down to scrutinise particular aspects of our research, given the broad spectrum of gamification approaches available for recruitment and selection, as unveiled during our exploratory research. As an outcome, we delve into and elucidate the emergent notion of **Gamified Situational Judgement Test (GSJT)**, including its design and manufacturing processes.
- In the second phase of the research, our attention pivots towards formulating proposals that specifically address the hurdles involved in assisting multiple stakeholders in the design and conception of Gamified Situational Judgement Tests. Initially, these proposals emerge as **a meta-framework** to streamline efficient collaboration among design stakeholders. This strategy accentuates the gamification process and offers **two initial frameworks and a modular design structure model** for crafting in-game situations. We also introduce a **drama manager** designed to facilitate the automatic generation of unique GSJTs, thus augmenting their **replayability and ensuring a diverse array of scenarios** for users to engage with. In addition, we also propose an approach to aid psychometric experts in initialising the GSJT scoring system via a **computational psychometric scoring system**. We

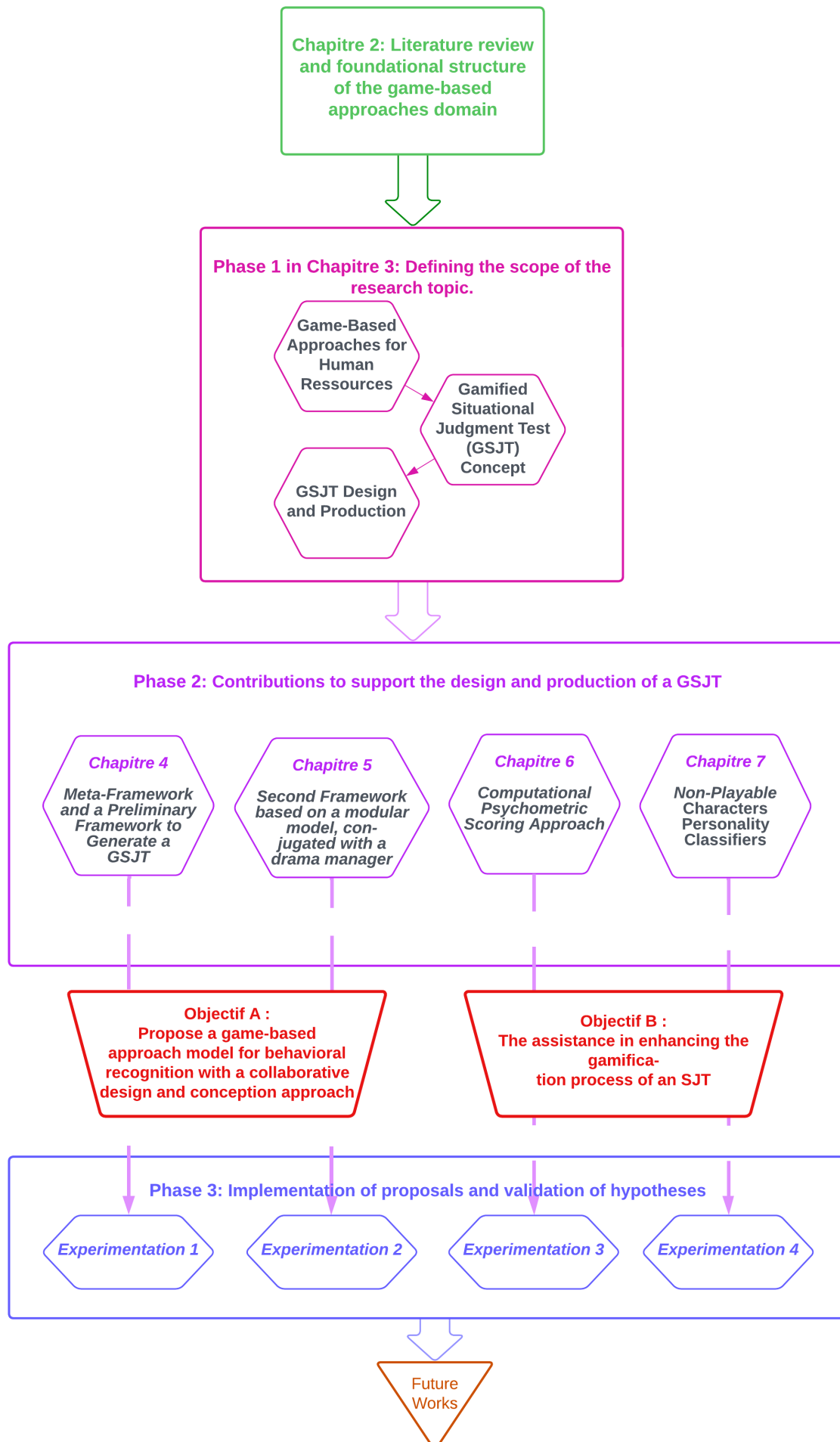


FIGURE 1.2: Our methodological approach

further propose to assist writers in crafting character dialogues, accurately reflecting realistic personalities based on **binary classifiers**.

- Lastly, in the third and final phase of the thesis, we delineate the outcome of our experimental procedures, with the objective of answering our research questions and goals. In pursuit of *Objective A*, which centres around proposing strategies to generate a game-based assessment for behavioural recognition efficiently, we implement our frameworks and evaluate them using two Gamified Situational Judgement Test prototypes: **an initial GSJT** based on the first framework, which subsequently undergoes revisions and improvements, and a **second GSJT** grounded on our second framework and the implementation of a dynamic modular system. Concurrently, in line with *Objective B*, which focuses on exploring the facilitation of stakeholders in enhancing the gamification process of a traditional situational judgement test used in recruitment and selection methods, we implement our innovative ‘computational psychometric scoring system to assess participants’ responses in an experiment based on the refined first GSJT. Besides the GSJT, we also construct and appraise our character personality classifiers. We scrutinise the dialogues of various game characters and employ our classifiers to assign personality traits to each individual character.

Chapter Overview

This chapter introduces our research context, focusing on **Soft skill and behavioural recognition for selection and recruitment purposes**.

The chapter established that soft skills encompass traits such as **personality attributes, interpersonal and social proficiencies, and cognitive capabilities, among various other**. It then examined their significance and the imperative to identify behavioural profiles for recruitment and selection purposes. The topic of how to measure soft skills was also addressed, touching upon various **psychometric models** used in their assessment.

Following this, the chapter delved into the specific aspects of our research subject, soft skill and behavioural recognition through **Game-Base Approaches**, while considering their financial implications for companies.

Hence, we defined our primary thesis's scope, direction and objectives by providing two specific objectives:

- **Objective A: Develop a game-based approach for immersive behavioural assessment**
- **Objective B: Propose innovative strategies to streamline the creation and gamification process.**

To address these objectives, we formulated the hypotheses based on preliminary studies of existing approaches in this chapter. Lastly, we detailed our research methodology.

Chapter 2

Game-based Approaches History

Game-based approaches have a long and rich history in various fields, including education, psychology, and human-computer interaction. The use of games for learning and training purposes has been documented as far back as the antiquity centuries, and the popularity of video games in the latter half of the century has led to a renewed interest in game-based approaches across many different domains.

In this chapter, we will explore **the history of game-based approaches** in depth, drawing on the work of scholars in the domain to provide a comprehensive overview of the evolution of this field over time. We will begin by examining the earliest uses of games for learning and training, including military simulations and board games used for **educational purposes**. We will then explore the rise of **video games and computer-based simulations** and their increasing use in areas such as healthcare, business, and the military.

Throughout this chapter, we will explore the fundamental theoretical and practical considerations pertaining to game-based approaches based on Deterding's study [52] and his conceptual map of game-based approaches (refer to Figure 2.1). We will delve into their efficacy for learning and training, the significance of game design and player engagement, and the diverse range of game-based approaches employed in different domains. By the conclusion of this chapter, readers will possess a comprehensive grasp of the historical development of game-based approaches, along with an understanding of the current landscape and potential future directions in this field.

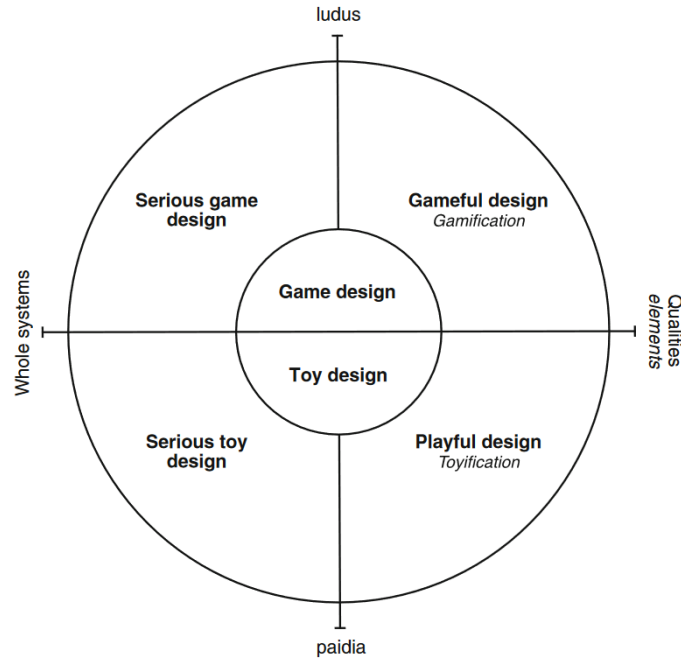


FIGURE 2.1: Conceptual map of Game-Based Approaches by Deterding et al. (2016)

2.1 Game-based Approaches' Origins

Throughout history, games and play have been used for more than just leisure activities but also as essential elements for **game-based approaches (GBA)**. According to *Herodotus' Histories*¹, games were invented by the *Lydians* as a distraction from a prolonged period of famine [53]. The games included dice, knucklebones, balls, and others. In Sophocles' plays, the Greek hero *Palamedes* is credited with creating the game of pawns during the Trojan War. The game was invented to alleviate the tension and conflict caused by a famine that threatened to tear apart the Achaean army. Hence, games were created to soothe stress and tension and maintain social community cohesion through entertainment. Entertainment often refers to activities such as games that provide enjoyment, pleasure, and amusement. It also refers to a way to engage and keep a person's thoughts.

Additionally, Krentz [54] reveals that in *Plato's Republic*², the play was highly related to the education of the citizens in establishing a better and just city. Krentz highlights that the Greek terms *paideia*, *paidia*, and *paides*, which respectively mean *education/culture*, *play/game/pastime/sport*, and *children*, share the same

¹Histories (Herodotus) - [https://en.wikipedia.org/wiki/Histories_\(Herodotus\)](https://en.wikipedia.org/wiki/Histories_(Herodotus))

²Republic (Plato) / [https://en.wikipedia.org/wiki/Republic_\(Plato\)](https://en.wikipedia.org/wiki/Republic_(Plato))

etymological root. As a result, these three terms are frequently used in the same context in Greek.

In another context, *play* and *game* concepts also show prominence in teaching behavioural and military skills. The Chinese emperor *Yao* invented the game of *Go* 2.2 to teach his son the virtues of discipline, concentration, and balance. The game is played by two players who take turns placing black and white stones on a grid with the aim of capturing their opponent's stones. Due to its strategic and tactical elements, *Go* gained popularity as a leisure activity among Chinese military leaders. However, from a leisure activity to a serious activity, the game revealed its potential as a training tool in the art of war. The classic military treatise, *The Art of War* by Sun Tzu [55], highlights the applicability of *Go*'s principles to military strategy. As such, the game served as a platform for training in tactical thinking, deception, and psychological warfare [56].



FIGURE 2.2: Game of Go

Likewise, in the 19th century, the Prussian military strategist, *General Carl von Clausewitz*, was often credited with laying the groundwork for using game-based approaches as training tools for military purposes. Clausewitz emphasized the importance of understanding the enemy's strategic objectives and the need for a structured, analytical approach to military planning [57]. Therefore, in the early

19th century, the Prussian army, following the works of Reisswitz Sr. and Jr., began developing a game-based approach that may be referred to as a **simulation game** that simulated military scenarios in order to train their officers in strategic planning and decision-making [58]. This simulation game was called *Kriegsspiel*, which means *war game*. The game provides a realistic simulation of the complexities of battle and teaches officers how to analyze and respond to different strategic situations. The popularity of war games quickly spread from the Prussian army to other military academies across Europe. By the late 19th century, many military academies used war games as essential to their **training programs**. These games were played using tabletop maps, specific miniature soldiers, and complex rule sets and often involved multiple players taking on different roles within a simulated military campaign (see Figure 2.3). Such use and design of physical elements may be called **serious toy design**. Indeed, serious toy design [52] refers to the process of designing toys with a specific purpose beyond just entertainment. These toys often aim to educate, develop certain skills, or address specific challenges or needs [59]. In the context of war games, miniatures are often meticulously designed to replicate real-world units, vehicles, terrains, and equipment. This level of detail aids in providing an authentic representation of warfare, which can be crucial for historical accuracy or strategic training. Finally, the impact of war games on military strategy and planning was significant, with many military leaders citing their value in improving decision-making and developing strategic thinking [60].



FIGURE 2.3: A Kriegsspiel Match in Progress (Photo by Kurzon - Wikipedia)

The use of war games continued to evolve in the 20th century as new technologies allowed for the development of more advanced simulations. In particular, the development of computer technology in the mid-20th century led to the creation of

computer-based war games, which allowed for even more realistic simulations of military scenarios. Military academies and training centres worldwide use it to prepare their officers for various **real-life scenarios**. Indeed, Djaouti et al. [61] discuss the development of computer-based war games for military training purposes, explicitly highlighting the example of *Hutspiel*. *Hutspiel* was a highly detailed war game created in 1955 by the *Operations Research Office at Johns Hopkins University*, which allowed players to simulate the impact of nuclear weapons on a global battlefield. Several other war games, including *T.E.M.P.E.R.*, a cold war simulation game, as well as the game *AGILE*, designed to simulate insurgency and counterinsurgency operations, were created in the 1950s and 60s, mainly by the *Joint War Games Agency* and a team led by Clark Abt.

2.2 Introduction to Serious Games

Clark Abt [62] is well-known for his contributions to simulation games for military purposes and his research that laid the foundations for using games as tools for **serious purposes** such as education, training, public policy, and social change. Indeed, Clark Abt proposed and defined the term *serious games* as:

- *Games that may be played seriously or casually. We are concerned with serious games in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement. This does not mean that serious games are not, or should not be, entertaining.*

Abt's definition puts any simulation game under the umbrella of serious games. However, several definitions have been proposed in the scientific literature, showing a diversity of concepts used to describe the term *serious games*. One of the more general definitions comes from Chen and Michael [3], who describe serious games as *games whose first purpose was not mere entertainment*.

Zyda [63] proposes another more specific definition that defines serious games as *a mental contest, played with a computer following specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives*. This last definition incorporates

the use of computer machines. It specifies that serious games are mental contests rather than mere games. Sawyer [64], the founder of *The Serious Game Initiative* in 2002, an independent institution aimed at developing Serious Games and the industry surrounding it, has also put forward a definition that aligns with the use of technology and video games outside of entertainment. Alvarez and Djaouti [65] also proposed a similar definition: *A serious Game is a computer application, for which the original intention is to combine with consistency, both serious (Serious) aspects such as non-exhaustive and non-exclusive, teaching, learning, communication, or the information, with playful springs from the video game (Game)*. While different definitions exist, they all share the common thread of games being used as a tool to achieve **serious objectives**. Ultimately, serious games have great potential to engage learners and affect meaningful change in various fields. In conclusion, serious games can include various types of games that may not only be digital and used for serious purposes such as education, training, public policy, or social change.

2.3 Serious Games are not only Digital

Alvarez and Djaouti [65], revealed that Kevin Corti point of view, the CEO of *PIXELEarning*³, may not agree with the narrow definition that serious games are only using computer technologies. Others agreed with the fact that serious games are not only digital. For example, *The New Alexandria Simulation: A Serious Game of State and Local Politics* is a simulation game designed to help teach students about state and local politics [66]. The game is based on the idea that in-game cities are centres of learning, culture, and innovation. In the game, students take on roles such as mayor, city council member, or business owner and work together to make decisions that affect the city's growth and development. The game includes a range of scenarios, from economic development and environmental issues to social justice and public safety. The simulation is designed to be highly **interactive and engaging**, with players learning about the complexities of real-world political issues and the challenges of working together to make decisions that benefit the community. Students gain a deeper understanding of the political process and how local government works by participating in the game.

³PIXELEarning is specialized in the design and development of serious games and immersive learning solutions

Nevertheless, it is a simulation made in real life without any technology. Other studies revealed that serious games that do not need to rely heavily on technology, known as offline, facilitated game-based experience. These offline games offer a range of benefits, including increased social interaction, lower costs, and the ability to reach audiences with limited access to technology. For example, *Fishbanks* is a role-playing game designed to teach players about managing renewable resources such as fish stocks [67]. Players take on the roles of fishing companies and make decisions about fishing quotas and investments in new technology. Studies have shown that playing *Fishbanks* can improve players' understanding of sustainable resource management and their ability to negotiate solutions to complex problems [68, 69]. Serious games using cards or board game setups that can be played offline are also studied for their cognitive and affective learning outcomes. *EcoChains: Arctic Crisis* is a multiplayer card game that focuses on teaching the impacts of climate change on Arctic ecosystems, specifically the diminishing sea ice caused by climate change. In the game, players must build food webs that are resilient to changing environmental conditions while meeting the needs of the human population [70].

In light of the various interpretations surrounding the concept of serious games, it is evident that any singular definition of the term inevitably entails certain limitations. Nonetheless, we must adopt a specific position to progress in our discourse on this topic. Therefore, we have chosen **to define serious games as encompassing all types of games that have the potential to serve serious objectives such as behavioural training and behavioural recognition**. This definition is intentionally broad, as it acknowledges the vast array of possible applications of serious games in many contexts, such as non-exhaustive and non-exclusive education and training in healthcare and rehabilitation. Ultimately, this thesis contributes to the ongoing academic discussion surrounding serious games, and we hope this definition will facilitate further exploration and innovation in this field. Nevertheless, we will focus the studies on **digital game-based approaches** for this doctoral thesis in computer science.

2.4 Simulation Games are Serious Games

Following the thesis position about the serious game definition, **simulation games**, as briefly explained earlier, are serious games and may be considered as a genre

of the serious game [4]. One of the most famous simulation games that represent well that a simulation is a serious game is the *Microsoft Flight Simulator*⁴ / *Microsoft*. This popular flight simulation game developed by *Microsoft* in 1982 has gained popularity since its creation and has been updated, improved, and is still used and played nowadays. The game allows players to simulate flying various aircraft in a highly realistic and detailed virtual world, including commercial airliners, small planes, and even helicopters. The game uses satellite imagery and data from *Bing Maps* to create a realistic 3D world with accurate terrain, weather, and other real-world conditions. Thanks to its realism, it has been used for **pilot training and education and for entertainment purposes** [71, 72]. Hence, such simulators revealed that even a game might be used for serious purposes and entertainment. In addition, scientific research also used the game to study the effects of climate change on aviation. Indeed, understanding its capabilities and limitations to effectively use such simulation games for training, safety investigation, and engineering and scientific research is needed [73]. However, simulation games are only one type that falls under the scope of our thesis' serious game definition. With the democratization of the use of computers and the advancement of technologies in the video games industry, games with serious purposes, such as edutainment software and games, were well risen through the 1980s and 1990s.

2.5 Edutainment Games, the Game-Based Learning Precursors

Edutainment games are serious games designed to combine entertainment and educational content and implemented in various fields, including this non-exhaustive list: education, health, and business. Edutainment games positively impact learning outcomes, providing an engaging and interactive environment that promotes active learning and knowledge retention [74]. One of the earliest examples of serious games for education was *The Oregon Trail*⁵, a classic educational game series that was first developed in the 1970s and continues to be used in classrooms today. *The Oregon Trail* 2.4 uses a role-playing game format to teach students about the challenges of frontier life and Westward expansion in the 19th century [75].

⁴Microsoft Flight Simulator

⁵The Oregon Trail / Minnesota Educational Computing Consortium

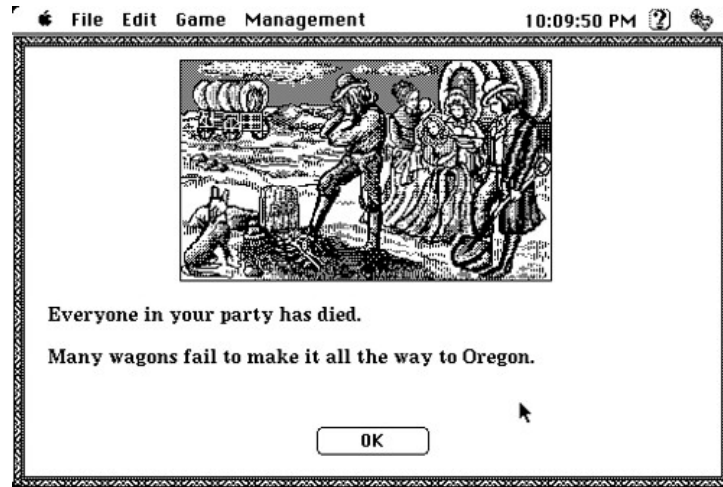


FIGURE 2.4: Oregon Trail Game

According to Moreno-Ger et al., [76], edutainment games can provide an effective means for teaching complex concepts and developing skills such as problem-solving and critical thinking. For example, various examples of edutainment games, including *Math Blaster*, *Reader Rabbit*, and *Where in the World is Carmen Sandiego?* have been famous in education and have been found to improve learning outcomes and **hard and soft skill development** [77]. *Math Blaster* and *Reader Rabbit* are two classic examples of educational video games used for decades to help children develop foundational skills in math, reading, and other subjects [78, 79]. *Math Blaster* (see Figure 2.5 is a series of games first released in the 1980s and designed to make learning math fun and engaging for children. The game features a science-fiction setting and tasks players with solving math problems to progress through levels and defeat alien enemies. Research has shown that educational games like *Math Blaster* can be effective in helping children develop math skills [80]. For example, a study by Luckhardt Redfield et al. [81] found that children who played an educational math game similar to *Math Blaster* may improve their math skills significantly. *Reader Rabbit* (see Figure 2.5, on the other hand, is a series of educational games designed to help children improve their reading and language skills. The game features a variety of activities, such as word recognition, spelling, and vocabulary development, all presented in a colourful and engaging format. Research has shown that educational games like *Reader Rabbit* can improve children's reading skills. For example, a study by Mendez et al. [82] found that children who played an educational game similar to *Reader Rabbit* may demonstrate significant improvements in oral reading skills.

However, serious digital games may also train cognitive ability and soft skills training purposes. For example, in their studies, Medalia et al. [83] use the edutainment game *Where in the World is Carmen Sandiego?* (see Figure 2.5 to highlight the positive result in treating problem-solving deficits in acutely ill psychiatric inpatients. Likewise, Carroll and Terry [84] promote the approach of using the edutainment game as a motivation tool to learn and introduces to a range of subjects, such as history, mathematics, language arts, science, music, and fine art, as well as it is capable of developing a variety of critical skills, including critical thinking, comprehension, writing, vocabulary, and research. Hence, the game may be used as a versatile and effective learning tool. Furthermore, the game's ability to foster collaboration, problem-solving, and decision-making through group learning processes and sessions can contribute to students' social and emotional development. Despite the wide range of possibilities the serious game offers, De Lisi et al. [85] reveal in their research that using *Where in the World is Carmen Sandiego?* to improve children's mental rotation accuracy may not be appropriate compared to the non-serious but famous game called *Tetris*. Nevertheless, their research highlights one of the aspects in the definition of Abt's about games, which may be any game that may be played seriously for serious purposes, but they are not serious games. Alvarez et Djaouti [65] proposed to call such use of *Tetris* a serious gaming approach.

2.6 Serious Gaming and Serious Game

In line with the discourse on serious games, it is noteworthy that a video game initially designed solely for entertainment purposes can be repurposed for educational or other serious pursuits. By adopting a serious posture toward gameplay, individuals can harness the benefits of video games beyond mere entertainment and unlock their full potential as a tool for learning and personal development. This phenomenon is not limited to a particular industry or sector; numerous examples abound. In the early 2000s, scholars have documented instances where video games have been leveraged in educational contexts to promote engagement, active learning, and skill acquisition [86, 87]. In his seminal work, Gee [87] offers compelling insights into the transformative potential of video games in education. Gee's research sheds light on how children who engage with historically and mythologically inspired games such as *Age of Mythology* and *Age of Empires*



Math Blaster Game

Where in the World is Carmen Sandiego? Game



Reader Rabbits Game

FIGURE 2.5: Math Blaster, Where in the World is Carmen Sandiego? and Reader Rabbits Games

become active learners, delving deep into the subject matter within and beyond the game. Specifically, Gee observes that these young learners develop an insatiable curiosity about mythology, often seeking out additional sources of knowledge such as websites and library books to supplement their understanding of the subject. Moreover, these learners begin **to see connections between their games and other domains** of popular cultures, such as superhero stories and even the broader arc of history and society. This mode of inquiry, facilitated by gameplay, allows individuals to deeply engage with the complex subject matter in an entertaining and impactful way, underscoring the rich potential of video games as a tool for education and personal growth.

During the same period, massively multiplayer online role-playing games became a

real trend that enables scholars to study the learning potential for the players these games may offer. A massively multiplayer online role-playing game (MMORPG) is a type of online game where many players interact with each other in a virtual world, taking on fictional characters and engaging in various activities such as combat, exploration, and socialization. In the scientific literature, Peterson [88], Rankin et al. [89], and Kongmee et al. [90] have employed massively multiplayer online role-playing games (MMORPGs) to explore the potential benefits of learning a second language through gameplay. These researchers have identified the potential of MMORPGs in addressing challenges commonly faced by language learners in real-life settings, such as providing opportunities for practice, feedback, and interaction with native speakers. Their research has discovered that MMORPGs can be highly effective in supporting language learning, as evidenced by improvements in standardized language tests and students' increased participation and progression in the game itself. Specifically, students who played MMORPGs exhibited increased fluency and confidence in using English, demonstrated tremendous patience when reading and writing, and were more motivated to produce dialogue when speaking and chatting. Moreover, multiplayer games can be a powerful tool for teachers to provide feedback and support in a friendly and interactive manner, fostering a **positive learning environment** [91]. In multiplayer games, learning takes place through conversation characterized by spontaneity, contextuality, and unpredictability. Playful and lighthearted emotions emphasize this type of learning, which can create an engaging and enjoyable learning experience. Through cooperative learning facilitated by the game, players from diverse cultural backgrounds can interact and learn from one another. The learning process in multiplayer games is multifaceted. It can support a variety of learning objectives, such as simple fact-finding, the development of tactics and strategies, and the working out of a moral order [92]. By conversing with other players, individuals can access a wealth of knowledge and experience that may not be available through traditional education. Furthermore, the game's immersive and interactive environment can provide a safe space for players to experiment with new ideas and approaches. Serious gaming, or the repurposing of video games for educational or serious purposes, has been exhaustively documented since the early 2000s to promote engagement, active learning, and skill acquisition. While serious gaming differs from a serious game, serious gaming has preceded the advent of digital game-based learning and has shown great potential for education and personal growth.

2.7 Digital Game-Based Learning

Following the research of Anastasiadis et al. [93], **digital game-based learning** is a pedagogical approach that integrates games into the educational process. By leveraging the motivational psychology behind games, this method fosters student engagement with educational material and subjects in a dynamic, enjoyable, and playful way. To further promote motivation and engagement, Prensky [94] introduced digital game-based learning as a student-centred approach that integrates digital games with curricular content. The essential characteristic of this approach, as defined by Prensky, is the convergence of interactive entertainment and serious learning through digital games. Erhel and Jamet [95] also promote the use of digital game-based learning to support educational purposes such as tracking and learning and characterize it as a competitive activity that sets educational goals to promote students' knowledge acquisition. Games and simulations that promote cognitive and soft skills allow students to practice these skills in a virtual and safe environment. Cheng et al. [96] also argue that digital game-based learning combines educational content with digital games to stimulate student interest and enhance learning effectiveness, which can lead to students viewing knowledge and education positively throughout their lives. Finally, Perrotta et al. [97] identified intrinsic motivation, learning through intense enjoyment and fun, authenticity, self-reliance and autonomy, and experiential learning as the fundamental principles involved in digital game-based learning. This research provides strong evidence for the potential of digital game-based learning as an innovative and practical approach to teaching and learning, with the potential to positively impact students' attitudes toward education and knowledge acquisition. According to Spires [98], game-based learning is not only about developing games for students to play but also about **designing interactive learning activities that gradually convey concepts and guide students toward specific learning outcomes. Hence, in order to maximize the benefits of digital game-based learning and provide students with a more comprehensive and enriched educational experience**, it can be integrated with other established learning methodologies and approaches such as cooperative learning [99], project-based learning [100], and problem-based learning [101]. By combining these different techniques, educators can create a more dynamic and immersive learning environment that engages students and helps them to develop a wide range of skills, including problem-solving, critical thinking, collaboration, and communication [93]. Hence, this suggests

that digital game-based learning pertains to the process, activity, or strategy employed to attain education objectives rather than the final product it produces. Finally, by expanding the objectives beyond education, start-ups, think tanks and digital agencies led in the 2010s to the emergence of the pairwise terms *gameful design/gamification* and *playful design/toyification*. Brought under the umbrella of *play/game-inspired design*, *gamification* and *toyification* aimed to apply and define the process of using games and play on a larger scale [102].

2.8 Play and Game Inspired Design: Overview

Game and play-inspired design is an approach to design that takes inspiration from games and play to create products, services, or experiences that are not only functional but also enjoyable and motivating [103]. This approach recognizes the power of play and games to engage people and create a sense of flow [104, 105], and seeks to apply these principles to a wide range of contexts, from business and education to health and wellness [106].

Playful and gameful concepts are the foundation of play and game-inspired design. Despite their similarities in the need for a safe environment for a player, the terms playful and gameful differ in meanings. Roger Caillois [107] made a seminal distinction between the concepts of play and games, introducing the terms *paidia* and *ludus* to describe two different forms of play. *Paidia*, or free-form play, is characterized by **improvisation, expressiveness, spontaneity, and joy** and is commonly observed in children's play. In contrast, *ludus*, or formal play, is governed by **rules and arbitrary obstacles** and often involves competition with defined winners and losers. This type of play is commonly found in board games and video games. Caillois' definitions place *paidia* (playful) and *ludus* (gameful) at opposite ends of a play continuum, highlighting their contrasting qualities and the diverse experiences they offer. Lucero et al., [108] proposed that research about playful and gameful experiences' pros and cons may enhance game-based approaches in whatever domain they are used.

One of the key tenets of game-inspired design is the focus on creating enjoyable and meaningful experiences. This approach recognizes that people are more likely to engage with and retain information when having fun. For example, game-inspired design may create educational experiences that teach complex concepts in a fun

and engaging way [109], as well as create games that use the concept of playfulness as a way to address real-world problems, such as climate change or social justice [106]. Playfulness is when people approach even ordinary activities with a spontaneous and joyful attitude, characterized by a lack of seriousness, specific objectives, or real-world consequences [107]. That play can provide a powerful way to motivate and inspire people [110]. Playful experiences arise when individuals adopt a playful approach to their activities or playfully view the world. Examples of playful experiences may include carefree leaping into piles of fallen leaves during autumn, aimlessly swiping through the home screens on our smartphones, or mischievously drawing on the hood of a dirty car using a finger [108]. These experiences can be delightful and motivating. Another example of using playfulness to motivate people is the piano stairs at the *Odenplan Metro Station* in Stockholm, which encourage individuals to use the stairs instead of the escalator as each step was playing a musical note when stepped on by someone [111].

Game and play-inspired design approaches have evolved from various fields, such as educational technology and psychology. As such, it draws on a wide range of theories from these fields to explain the effectiveness of such approaches in achieving non-entertaining outcomes. One of the most influential theories in the field is the self-determination theory (SDT) [112]. SDT posits that intrinsic motivation, or the drive to engage in an activity for its own sake, is a key factor in sustained behaviour change. Games promote intrinsic motivation by offering players autonomy, competence, and relatedness, which are the basic psychological needs according to SDT [113–115]. Therefore, applying games' theories to a non-entertainment context may enhance the experience and achieve specific goals. Game and play-inspired design also emphasize the importance of player agency and choice, recognizing that people are more likely to engage with experiences that allow them to take control and make decisions [116]. This approach involves creating environments and systems that allow players to explore, experiment, and make decisions, leading to a sense of ownership and investment in the experience [117]. For example, a game or play-inspired design for education can offer students autonomy by providing them with choices in how they learn, such as the activities they engage in or the order of topics they explore. The game or play-inspired design can also give students a sense of competence by offering them challenges that are well-matched to their skill level, so they feel a sense of accomplishment as they progress. This later statement is based on another influential theory, which is the flow theory previously cited [118, 119]. The flow theory in game-based approaches

suggests that balancing challenge and skill characterizes optimal experiences. In-game and play-inspired design, this balance is often achieved through adaptive difficulty, where the play or game-inspired design adjusts the level of challenge based on the player's performance. For example, a game or play-inspired design that teaches a new language can use adaptive difficulty to ensure that players encounter new vocabulary words and grammar rules at a pace that is challenging but not overwhelming.

Overall, game-inspired design represents a powerful approach to creating products, services, or experiences that are both functional and enjoyable. This approach has been used to create various innovative solutions in various contexts and can potentially drive significant impact and value in the future [2, 64, 87, 120]. Incorporating these and other relevant theoretical foundations into the play and game-inspired design can help ensure the experiences are compelling, engaging, and well-suited to their intended purpose. Therefore, these approaches have gained popularity in recent years, as they offer a unique and engaging way to tackle a wide range of challenges in almost any context. Hence, experienced designers have been brought in to create many game concepts and play-inspired designs, including gamification and toyification. Although these concepts are related to the use of game and play elements and mechanics in non-game contexts, they have some essential differences that we will define in the following subparts.

2.9 Gamification and Toyification

Gamification and toyification are concepts often used interchangeably in popular discourse, but they represent distinct approaches to incorporating game elements into non-game contexts [2]. While both gamification and toyification involve using play and game elements in non-game contexts, as explained previously, they differ in their underlying goals, design principles, and outcomes. For this thesis, we defined the term gamification as the process of **incorporating game elements, such as points, storytelling, badges, and leaderboards, into non-game contexts**, such as education, health, and business [2]. Gamification aims to motivate users to engage in desired behaviours by making them more enjoyable, rewarding, and meaningful. Gamification draws on principles of behavioural psychology, such as operant conditioning and self-determination theory, to design game mechanics that appeal to users' intrinsic and extrinsic motivations[121, 122].

Furthermore, gamification outcomes typically focus on specific behaviours or tasks, such as completing a training program, adopting a healthy habit, or buying a product [2]. Gamification outcomes, such as increased user engagement, improved performance, and higher sales conversion rates, are often measurable and tangible.

Toyification, on the other hand, is defined for this thesis as **incorporating play elements, such as imagination, exploration, and experimentation, into non-game contexts** [2]. The goal of toyification is to stimulate users' creativity, curiosity, and sense of wonder by creating a playful environment that encourages exploration and experimentation. Toyification draws on principles of play theory, such as Huizinga's [123] concept of *magic circle*, to design playful experiences that blur the boundaries between reality and fiction. In addition, toyification outcomes are often intangible and subjective, such as an increased sense of playfulness, enhanced imagination, or stronger social connections. Toyification outcomes are often difficult to quantify or measure, but they may have long-term and far-reaching effects on users' well-being and life satisfaction.

The design principles of gamification and toyification reflect their different goals and theoretical foundations. Gamification emphasizes using game mechanics that reinforce desired behaviours, such as rewards, feedback, and goals [124]. Gamification designers aim to create a funnel that guides users from initial engagement to sustained motivation, using techniques such as onboarding, scaffolding, and mastery [121]. In health and wellness, game-inspired design has been used and provided mediums that address real-world problems, such as obesity and physical inactivity. For example, *Zombies, Run!*⁶ is a popular mobile application that uses the format of a zombie apocalypse to encourage players to be more active and get in shape [125] (see Figure 2.6).

Game-inspired design has also been used to create new products and services that are both functional and enjoyable. For example, *Fitbit company* use gamification principles, such as points and leaderboards, to motivate users to be more active and healthy [126]. With their *Fitbit* (see Figure 2.7) devices, the company provides real-time feedback on users' activity levels and rewards and incentives for meeting fitness goals, making it a popular tool for promoting healthy habits.

Beale et al. [127] also studied using a game-inspired design approach to actively involve young people with cancer in their treatment using a game called *Re-Mission*

⁶Zombies, Run! / Six to Start

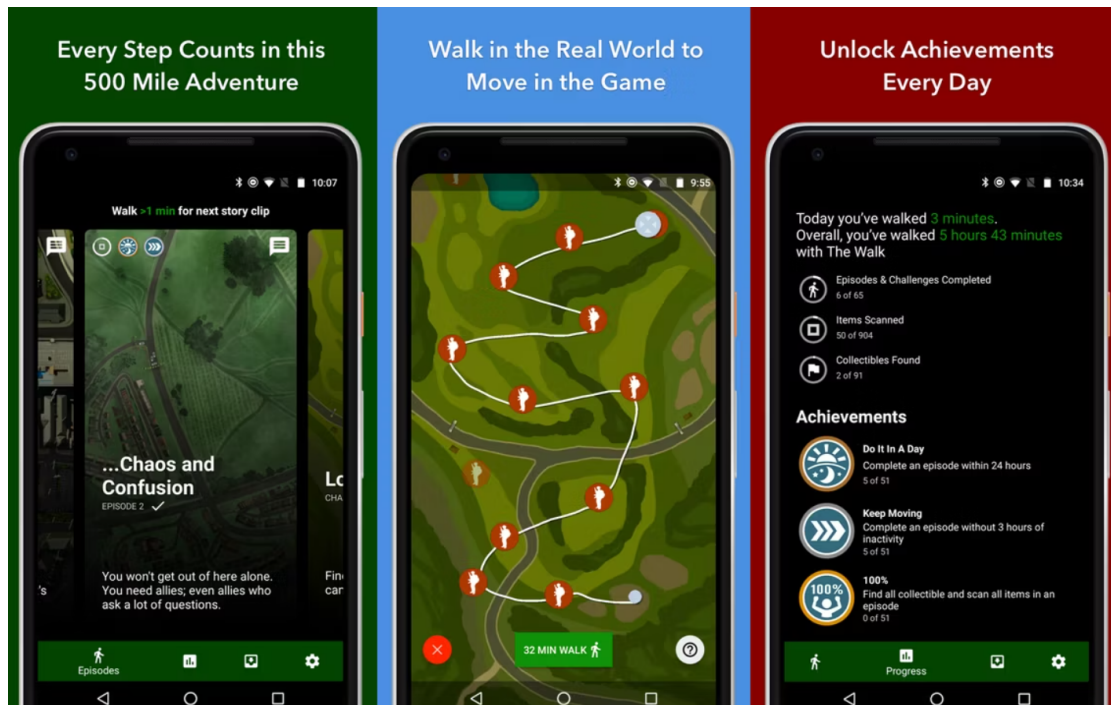


FIGURE 2.6: Zombies, Run! by Six to Start



FIGURE 2.7: Fitbit Gamification Example by Fitbit company

⁷ (see Figure 2.8). *Re-mission* is a game-based intervention aimed at improving the adherence of adolescent and young adult cancer patients to their medication regimens. The game was designed to be fun and engaging, providing players with feedback on their progress and tangible rewards for taking their medications.

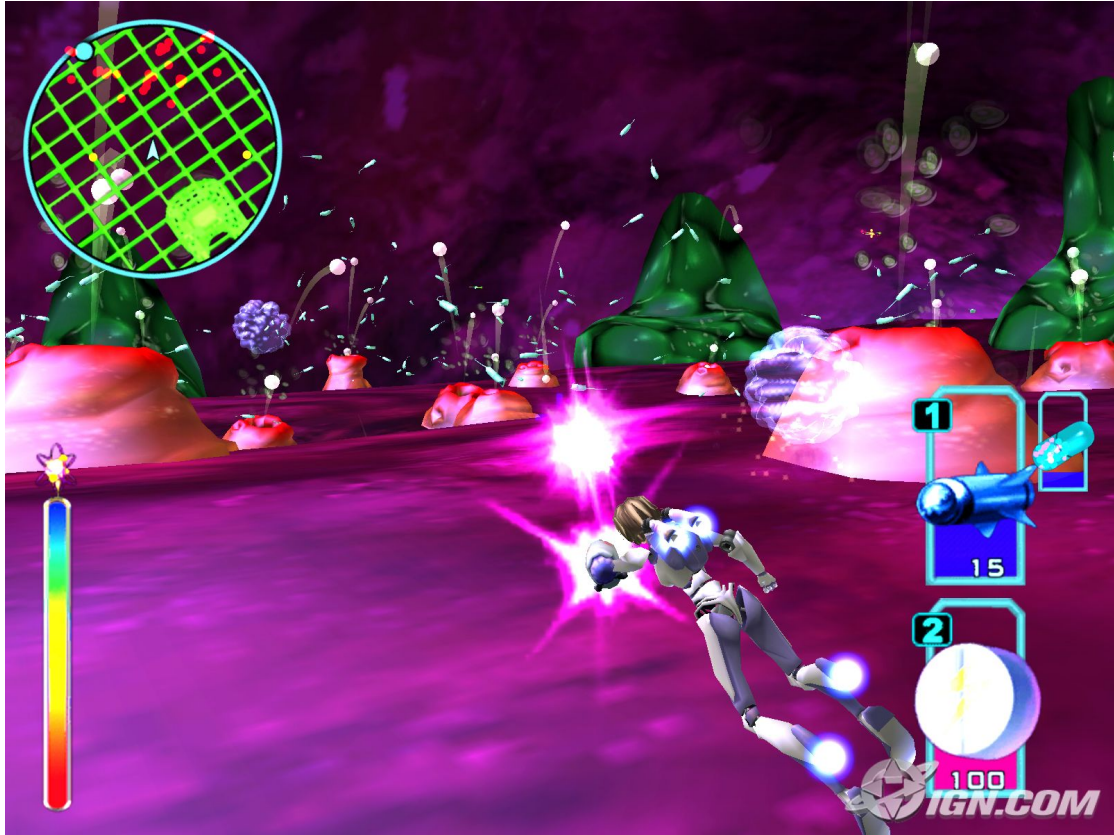


FIGURE 2.8: Re-Mission by HopeLab Foundation

Toyification, on the other hand, emphasizes the use of open-ended play experiences that stimulate users' curiosity and imagination [128]. Toyification designers aim to create a *sandbox* that allows users to explore and experiment with different possibilities using sensory stimulation, surprise, and whimsy techniques. Toyification designers often prioritize user autonomy and creativity over external rewards and constraints [129]. For example, in the business world, play-inspired design has been used to create *toyified* systems for employee training and motivation, such as *The Lego Serious Play* (see Figure 2.9) method⁸, which uses lego bricks and playfulness to facilitate creative problem-solving and teamwork [130].

⁷The idea for the Re-Mission games aimed at young cancer patients originated from Pam Omidyar, while the nonprofit organisation HopeLab Foundation conducted research to design them. The development process involved active collaboration with young cancer patients, oncology doctors, nurses, and game developer Real-time Associates, among other contributors.

⁸The Lego Serious Play / The Lego Group



FIGURE 2.9: The Lego Serious Play by The Lego Group

In the realm of education, play-inspired design has been used to create an educational and playful environment that teaches complex concepts fun and engagingly. For example, *Minecraft: Education Edition*⁹ is a popular educational and play-inspired platform (see Figure 2.10) that uses the format of sandbox-style games to teach subjects such as math, science, history, and more [131]. Another example of play-inspired design that has become increasingly popular in the business world, where it is used to create engaging and motivational employee training programs and customer engagement strategies, is the book *Gamestorming*. Despite its name that uses the term game in its title, the book provides a set of tools and techniques for using play in the workplace to solve problems and drive innovation [132]. The book also provides a range of play-inspired activities, from quick and simple games to more complex simulations, that can be used to create engaging and collaborative work environments.

In summary, gamification and toyification are two distinct approaches to incorporating game elements into a large panel of non-game contexts. While both gamification and toyification are approaches pinpointed under game and play-inspired

⁹Minecraft Education Edition is based on the original Minecraft sandbox-style game / Mojang AB. TM Microsoft Corporation.



FIGURE 2.10: Minecraft: Education Edition by Microsoft

design, they also share some similarities, such as the use of play and game elements. Nevertheless, they differ in their underlying goals and design principles.

2.10 Conclusion

In this chapter, we reviewed three game-based approaches, the serious game approach, the playful design approach and the gameful design approach, used in various domains and purposes, including education, health, and business. Through these examples, we have seen how games can be designed to leverage the motivational power of play to enhance user experiences, promote skill development, and facilitate behaviour change. For this thesis's objectives and original frame, we align our definitions of the four game-based approaches and definitions presented with a conceptual map 2.1 by Deterding [52], namely serious game design, playful design, gameful design, and serious toy design. The latter was briefly introduced in this chapter, as it does not fit the thesis' objectives of studying digital game-based approaches. Indeed, serious toy design focuses the design on a physical toy such as the *Fitbit* device or tabletop maps and miniature soldiers used in war games. In essence, these approaches, including serious toy design, are widely adopted in the academic and private fields of game-based approaches. They effectively achieve serious objectives, such as education, training, and behaviour change [6].

The first approach, serious game design, involves designing or deploying full-fledged games for serious purposes. Serious games can provide engaging and interactive experiences that motivate learners to achieve their learning goals. However, designing and developing serious games requires significant resources and expertise, which can challenge some organisations. The second approach, gamification, involves designing to afford the experiential and behavioural qualities characteristic of gaming. However, designing for gameful experiences requires a thorough understanding of game mechanics and user motivations, which can be challenging for some designers. The third approach, playful design, involves designing to afford the experiential and behavioural qualities characteristic of play. This approach creates non-game contexts that are more playful, creative, and exploratory for learners. The fourth approach, serious toy design (Toyification), involves designing or deploying full-fledged toys for serious purposes. To summarize, the four game-based approaches identified by Deterding and colleagues offer valuable insights and frameworks for designing and deploying game-based interventions for serious purposes.

In conclusion, using game-based approaches in various contexts has proven to be an effective and engaging way to achieve serious outcomes. The success of game-based approaches in education has been well documented, but the potential for these approaches goes beyond just the classroom. Game-based approaches have shown promise in areas such as health and wellness, workplace training, and even in military and defence. Furthermore, game-based approaches can be tailored to suit the specific needs of different contexts and audiences. By incorporating principles of gamification and game design, these approaches can enhance motivation, engagement, and learning outcomes. Using game-based approaches also offers a unique opportunity to collect data on user behaviour and performance, allowing for continuous improvement and optimization of the approach. By leveraging the power of gamification and game design, we can engage users in new and exciting ways and achieve outcomes that were previously difficult to attain. As such, the future of game-based approaches is promising. Hence, we will deeply explore the use of serious games and gamification for human resources to follow the original objectives and goal of the thesis, which are evolving and enhancing these approaches to impact various industries and fields.

Chapter Overview

In Chapter 2, we delved into the historical journey of game-based approaches. We begin by tracing the origins of these methods, transitioning into an introduction to the realm of **serious games**. It's important to note that Serious Games extend beyond digital platforms, and indeed, **simulation games** are classified as a subset of serious games.

The chapter discussed edutainment games, underlining their foundational role in the evolution of game-based learning. We also highlighted the nuanced differences between serious gaming and serious games.

Finally, we offered an overview of **play and game-inspired design**, highlighting the role of gamification and toyification in fostering user engagement and learning.

This historical exploration revealed that game-based approaches were predominantly used for learning and educational purposes. However, the application of such approaches to our contexts, which are **recruitment, selection, and behavioural recognition**, is a relatively recent development.

Chapter 3

Related Works

Human resources (HR) departments have always played a critical role in the success of an organisation. Historically, HR departments have been responsible for a wide range of activities that span the entire life cycle of an employee's involvement with an organisation. These activities include **posting job advertisements, screening and interviewing applicants, onboarding, benefits coordination, employee retention, performance evaluation, and exiting processes for retired or terminated employees.**

Over the years, HR departments have used various traditional methods to attract and select applicants. However, with the evolution of technology, the changing job market and the need to attract top talent, HR departments are exploring new and innovative ways to attract and select applicants. This change has led to the use of game-based approaches as innovative methods in the recruitment and selection process. This chapter aims to explore the use of **game-based approaches specifically for recruitment and selection purposes.**

In recent years, HR departments have been using game-based approaches (GBA) to create **interactive and engaging experiences** for applicants during the recruitment and selection process. GBA offers a unique way to **assess an applicant's cognitive and behavioural traits** that we will explore through this chapter, which traditional recruitment methods such as job advertisements may not capture.

This chapter will be structured as follows: In the first section, we will provide a brief overview of **the objectives and usage of game-based approaches**

for recruitment and selection purposes used by HR departments. In the second section, we will delve into **the concept of game-based assessment** and, more specifically, on **gamified situational judgement tests** based on interactive narrative games. Finally, in the last section, we will examine the challenges faced in **implementing a gamified situational judgement test** in HR practices.

3.1 Human Resources and Game-based Approaches for Recruitment and Selection Processes

In recent years, the human resources field has witnessed a surge in the adoption of game-based approaches to enhance various aspects of recruitment and selection processes [133, 134].

Despite GBA being utilised in HR recruitment and selection for several decades, there is a noticeable **lag in the academic literature's ability** to keep pace with the rapid changes occurring in practice. To address this discrepancy and provide a more comprehensive understanding of the field, Bina et al. [135] conducted a systematic review of GBA in the academic literature with a focus on HR recruitment and selection.

The review reveals that applying GBA in HR processes has the potential to enhance the effectiveness and efficiency of recruitment and selection significantly. Indeed, GBAs encompass a wide range of techniques, such as serious games, gamification, game-inspired designs, and serious toy design as seen in the Chapter 2, all aimed at achieving different objectives within the HR context recruitment and selection such as **enhancing candidate experiences or pool**, [136–139]. The growing interest in game-based approaches in the recruitment and selection process can be attributed to the increasing need for innovative solutions [140] that address the limitations of traditional methods [141]. Game-based assessments in recruitment and selection processes offer potential benefits for both **job applicants and organisations** [5, 142].

Several studies investigate the potential of gamification and serious games in recruitment and selection by drawing upon existing and prior scholarly work. For example, Armstrong et al. [133] suggest, on the one hand, that drawing upon

GBAs' principles, employee recruitment can be enhanced through two distinct approaches. Firstly, game-based assessments can be employed **to optimise the alignment between applicants and hiring organisations**. In essence, organisations that facilitate person-organisation fit with GBAs aid applicants in determining whether their values and expectations align with those of the recruiting organisation [143–145]. Robson et al. [146] support engaging employees through serious games and gamification to offer an opportunity to experience a virtual representation of the job and work environment before deciding on a candidate for a job.

Secondly, if organisations appear more appealing to potential candidates by **enhancing candidates' experience and improving employer branding in the overall recruitment process**, those organisations are likely to receive a higher volume and an overall better quality of job applications. Likewise, the findings of Gkorezis et al. [137] illustrated that using GBAs positively influences applicants' perceptions of an organisation's appeal. This enhanced perception, in turn, results in a heightened intention to recommend the organisation. Notably, this effect is particularly pronounced among the new generation of applicants with more extensive video gaming experience.

On the other hand, selection research has primarily concentrated on **the validity and reliability of selection predictors**. As suggested by Armstrong et al. [133] and Landers et al. [147], serious games and game-based assessment can provide a more **immersive and realistic environment** for candidates to demonstrate their skills and abilities, helping employers to better evaluate a candidate's fit for a role and potentially reducing the risk of bad hires. In addition, recent studies are also focused on striving for a **balance between minimising adverse impacts in fairness and equity and effectively forecasting job performance** while maintaining positive responses to the selection process [148].

Following and supporting the previous findings, the remainder of the subsections will detail the various uses of specific GBA, such as game-based assessment based on serious games and gamification in recruitment and selection processes.

3.1.1 Game-based Assessment to Enhanced Candidate Experience

Utilising game-based techniques such as serious games and gamification in recruitment processes not only elevates candidates' organisational perceptions but also influences their likelihood of job acceptance [137, 149], supported by increased assessment motivation [150].

Several empirical studies, including those by Landers et al. and Hamari et al., corroborate that these interactive assessments amplify candidate satisfaction and engagement [5, 151]. These assessments also cultivate a sense of procedural fairness, corroborated by research such as Leutner et al. [152], which posits that candidates perceive these methods as more equitable than traditional avenues. Additionally, the immediacy and depth of feedback in game-based assessments facilitate candidate self-reflection and performance understanding [153]. This feedback can also promote self-reflection, as candidates gain insights into their abilities and potential fit within the organisation [154, 155]. Moreover, game-based assessments can further be tailored to encapsulate specific job requirements, allowing candidates to align their skills more precisely with job expectations [154, 155].

Despite these advantages, the literature also highlights potential challenges and ethical concerns surrounding the implementation of this assessment, such as the risk of bias and discrimination, data privacy, and user consent [156]. In addition, some candidates may perceive game-based assessments as less credible than traditional methods, potentially affecting their experience [133]. Additionally, the cost and complexity of developing and implementing game-based assessments could be prohibitive for some organisations [157].

3.1.2 Game-based Assessment to Enhanced Candidate Pool

Leveraging the principles of the Attraction-Selection-Attrition framework [158], gamified recruitment practices have emerged as a potent tool for augmenting organisational allure, thereby diversifying the pool of prospective applicants. Case in point, L'Oréal's "Reveal," an online recruitment game (see Figure 3.1), serves the dual purpose of furnishing candidates with an interactive and realistic simulation of the workplace while concurrently granting L'Oréal a dynamic platform for

evaluating the candidates' aptitudes and potential contributions [159]. Within the game's simulated business setting, candidates assume various roles, make strategic decisions and manage a 'virtual' brand, thus acquiring experiential insights into L'Oréal's culture and operations. This innovative approach not only aids candidates in evaluating their fit with the company but also furnishes L'Oréal with valuable data on applicants' problem-solving and decision-making abilities.

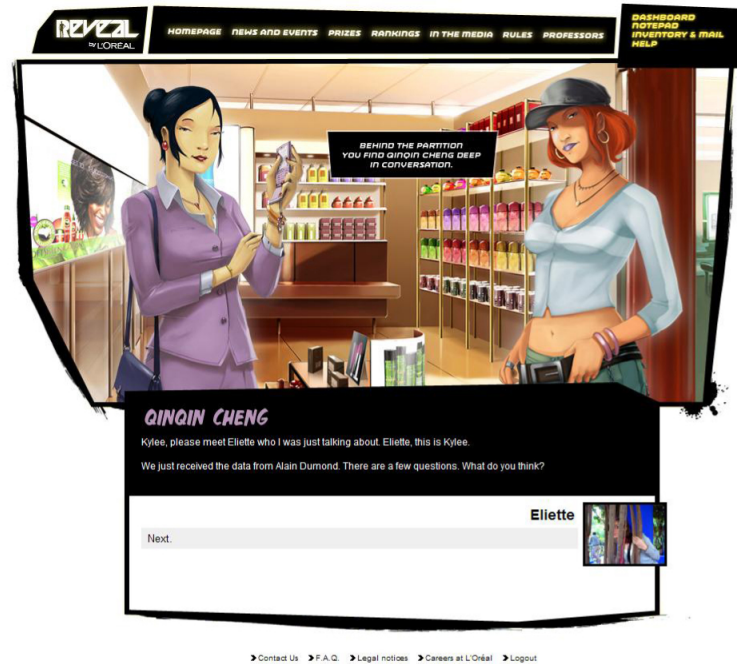


FIGURE 3.1: Reveal Serious Game by L'Oréal

Moreover, Joy's research [160] indicates that game-based assessments can effectively **identify high-potential candidates while improving the overall quality of the applicant pool**, mitigating biases related to gender and ethnicity in the process [152]. Such assessments also facilitate person-organisation fit by offering realistic job previews. Indeed, organisations could enhance person-organisation fit by offering **a realistic job preview** within the context of a serious game. A prime example of this approach is America's Army (see Figure 3.2), a game developed by the US Army to recruit young Americans. The game provides a glimpse into soldier life's thrilling aspects while showcasing army career opportunities and benefits. Players who perceive themselves as a good fit for the US Army based on their in-game success may be more likely to apply. Assuming that gameplay skills represent real-life social skills, this approach could yield a more qualified applicant pool than traditional methods.



FIGURE 3.2: America's Army: Proving Grounds by United States Army

For instance, Marriott Hotels' game-based assessment (see Figure 3.3), *My Marriott Motel* by Marriott Hotels [146], serves as an interactive tool for measuring both cognitive and non-cognitive skills of prospective candidates. By offering a simulated hotel work environment, the assessment not only enhances the recruitment process's efficiency by attracting applicants who possess a thorough understanding of the tasks and responsibilities in their desired positions but also attracts well-informed and role-compatible applicants. This innovation leads to better employee retention and satisfaction, outperforming traditional psychometric methods in engagement and fairness [146, 148].



FIGURE 3.3: My Marriott Motel by Marriott Hotels

While game-based assessments offer innovative and immersive recruitment experiences that can enlarge and enrich the applicant pool, they also risk misrepresenting an organisation's true character. For instance, a recruitment game could falsely project the organisation as environmentally responsible, attracting candidates with sustainability values [161, 162]. Such a mismatch could lead to increased attrition rates when employees discover the discrepancy between portrayal and reality. Despite potential drawbacks, GBAs enrich the candidate pool's quality and engage those disinterested in traditional recruitment methods. Moreover, they enable organisations to identify top talent comprehensively, thus improving workplace fit and overall applicant quality [49, 163].

3.1.3 Game-Based Assessment to Applicant Reactions to Selection Process

Applicants who view selection systems as unfair are likely to react negatively, impacting not only the hiring process by leading to heightened test anxiety and diminished test motivation [164] but also longer-term outcomes like job performance. This is explained by organisational justice theory, which suggests fairness perceptions are shaped by how rewards are allocated (distributive justice) and the rules behind them (procedural justice) [165]. Poor fairness perceptions can result in negative outcomes like lower self-esteem, reduced job offers, and higher turnover rates [164, 166]. Gilliland's [167] framework (see Fig 3.4) outlines how applicants interpret selection fairness, linking specific attributes of the process to perceived fairness.

Test motivation and organisational attraction are further moderated by computer proficiency. Indeed, technological experience, particularly in gaming, amplifies favourable responses to gamification. However, poor design can trigger perceptions of unfairness, especially in high-stakes settings, exposing organisations to potential legal risks if these perceptions correlate with protected class membership.

Indeed, recent studies indicate that procedural justice perceptions mediate the relationship between technology usage in selection and applicant outcomes like test motivation and organisational attraction [168, 169]. However, computer proficiency moderates the link between procedural justice and applicant outcomes,

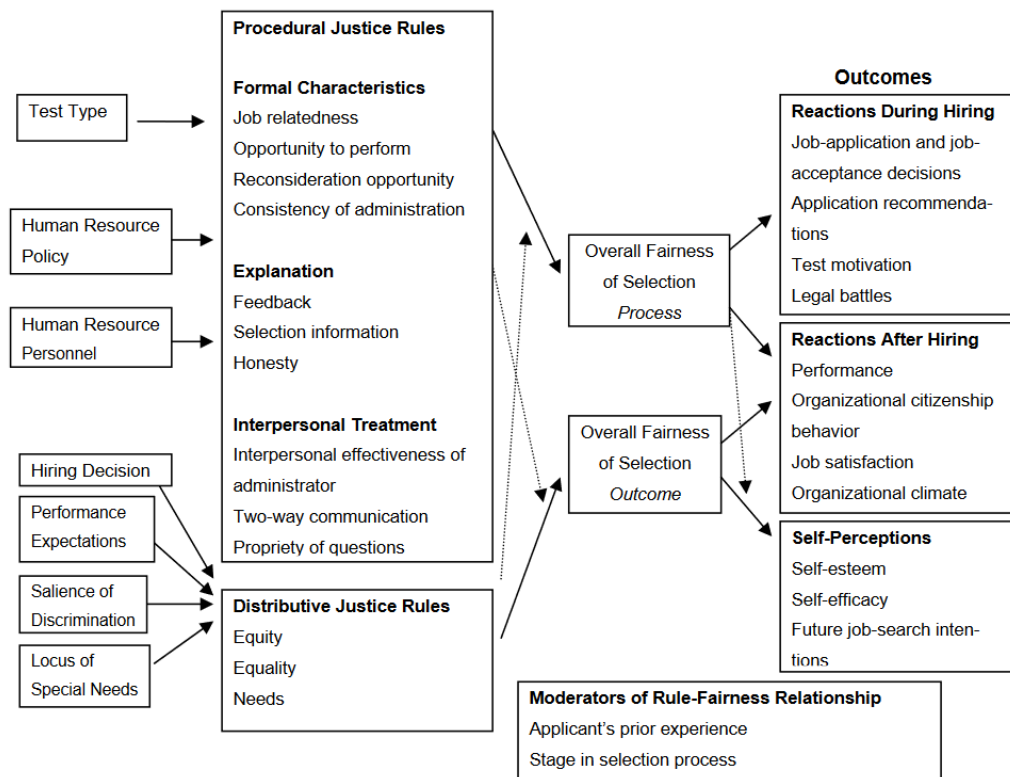


FIGURE 3.4: Model of Applicants' Reactions to Employment Selection Process and Outcome (from Gilliland, 1993, p. 700)

strengthening this association among those with greater tech experience. For example, prior gaming experience positively influences responses to gamification. [170]. In contrast, Landers and Armstrong also highlight that those with limited or no gaming experience may perceive the application of serious games in high-stakes contexts as inherently unfair if the game or gamified assessment is poorly designed or implemented [133]. The belief that a game is biased or unwinnable may contribute to the perception that the overall selection process is unfair, regardless of whether game performance accurately predicts job performance. If such perceptions differ according to protected class membership, the use of games in selection could potentially expose organisations to litigation risks [171].

In summary, well-designed game-based approaches in terms of fairness can enhance applicant reactions by promoting fairness and engagement, potentially leading to more accurate assessments and positive organisational perceptions.

3.1.4 Objective measurement in Game-Based Assessment

Armstrong et al. [133] suggest that game-based approaches and assessments objectively evaluate candidate abilities, as they may be designed to measure specific skills or competencies through **standardised scoring systems**. Moreover, GBAs have the potential to minimise applicant faking by offering engaging, immersive tasks that are challenging to deceive. In traditional recruitment methods like interviews, written tests, or self-report questionnaires, applicants may exaggerate their abilities or provide dishonest answers to present themselves favourably [172]. Hence, researchers have posited that GBAs may mitigate faking as they require participants to engage in complex tasks [133]. These tasks involve **situational judgement tests and scenario-based simulations**, which are more difficult to deceive than traditional methods. Furthermore, Landers and Sanchez [173], highlight three aspects of faking behaviour that designers of gameful assessments can target: motivation, ability, and opportunity (see Fig 3.5).

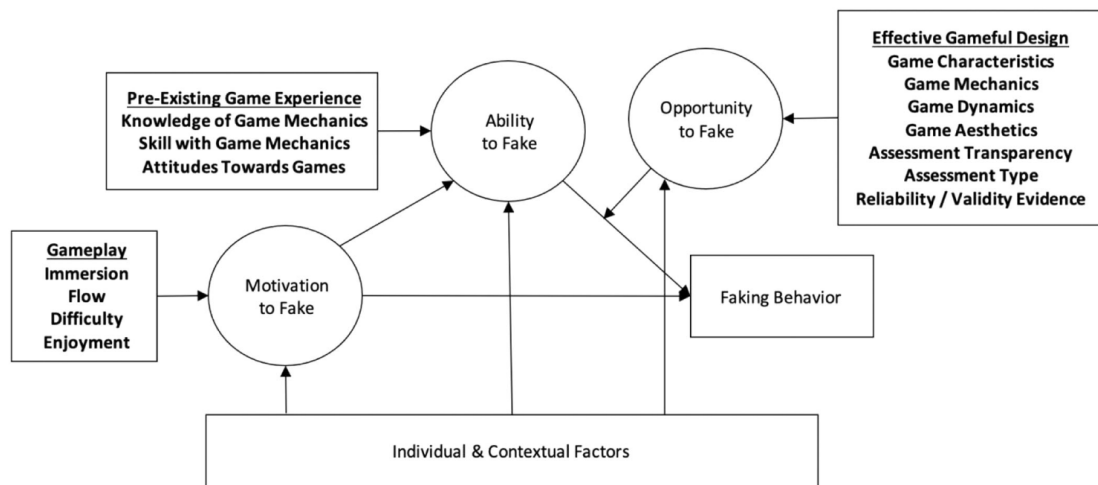


FIGURE 3.5: Suggested impacts of gaming elements on faking in game-based assessment. (from Landers and Sanchez, 2022, p. 9)

For example, by creating immersive experiences that induce a flow state, they can replace the motivation to fake with the motivation to play. To hinder the ability to fake, designers can use game mechanics that increase cognitive load, limit knowledge about effective faking behaviours, and capitalise on the novelty of game elements. Finally, designers can limit opportunities for faking through low assessment transparency, masking test purposes and desirable responses and using skill-based assessments that require authentic demonstrations of ability.

In addition, GBAs may be advantageous in providing objective measurement of soft skills, as they enable the collection of behavioural data that can be analysed quantitatively. A study by Stoeffler et al. [1] demonstrated that game-based assessments could successfully measure soft skills by analysing user interactions and performance within the game environment. Indeed, they designed *Circuit Runner* (see Figure 3.6), a game-based assessment where players navigate a 3D maze from a first-person perspective, working with a computer agent via a dialogue panel to complete challenges at locked gates. These challenges require information and resources asymmetrically distributed between the player and the computer agent. The chatbot is based on complex dialogue trees with prompts and multiple-choice response options designed to reflect different skill levels. Therefore, through the players' choices, the recruiters can track candidates' decision-making patterns, response times, and interactions with in-game characters or bots, which can be linked to specific soft skills. Unlike traditional self-report questionnaires, which rely on subjective interpretations and self-perception, serious game or gamified solutions may be designed to generate measurable, real-time data based on candidates' actions within the game, enabling a more objective understanding of their soft skills. However, a potential disadvantage of such an approach is the risk of misinterpreting behavioural data due to the complex nature of soft skills [174]. Potential issues may arise when relying solely on game-based assessments to measure candidate abilities. One concern is the validity of the game-based assessment, as it is essential to ensure that the game accurately measures the desired skills and competencies and does not inadvertently assess irrelevant factors.

Overall, while game-based assessments offer an objective way of evaluating candidate abilities, organisations must address the potential drawbacks and ensure that these tools are used alongside other selection methods to provide a comprehensive and fair evaluation of applicants[175].

3.1.5 Game-Based Assessment to Address Bias

Some game-based assessments based on serious games and gamified solutions have also emerged as a promising approach to **counteract biases** in the recruitment and selection process by offering several advantages such as addressing unconscious bias, reducing stereotype bias, minimising halo effect, enhancing objective decision-making, and encouraging diversity and inclusivity.

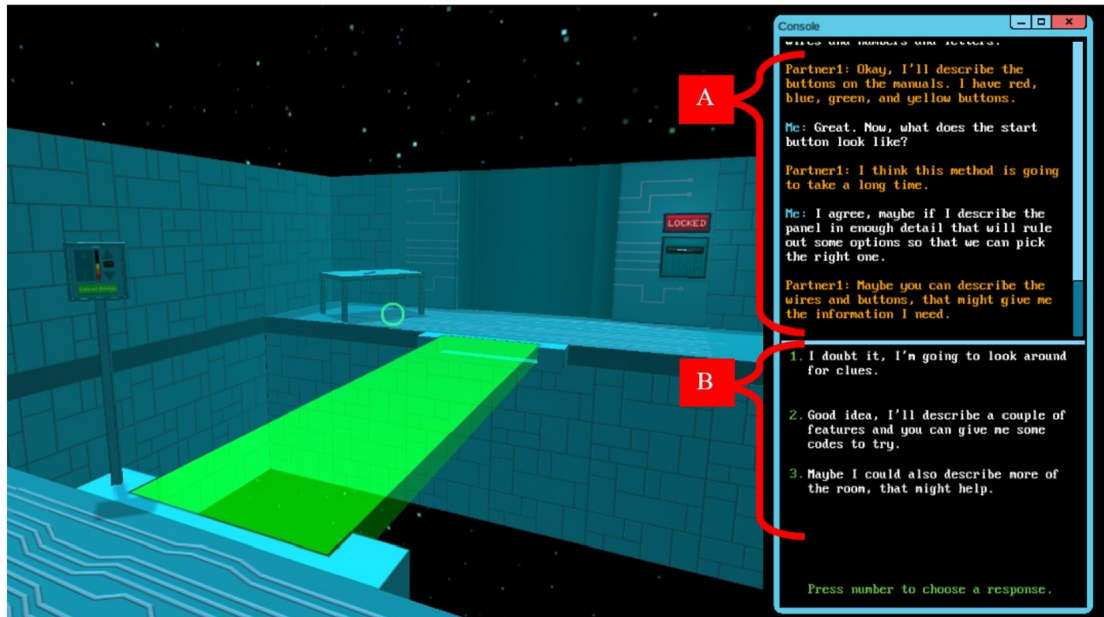


FIGURE 3.6: Circuit Runner game. The left division portrays the player’s navigation and interactivity within the labyrinth, whereas the right is subdivided into two segments. The upper subsection (A) showcases the conversational history between the player and the virtual agent. The lower subsection (B) presents the player’s possible responses and prompts a numeric selection of these choices.

[1]

Firstly, unconscious biases, such as stereotype biases, the innate human tendencies of automatic judgements and assumptions about individuals predicated on characteristics such as race, gender, and age, pose significant challenges in the recruitment and selection process [176]. For instance, game-based assessment, by using anonymised avatars or pseudonyms, can partially mitigate these biases by dissociating a candidate’s identity from their evaluation, thus focusing the assessment on in-game performance [133]. However, unconscious biases could still emerge based on nuances of candidates’ gameplay, such as decision-making speed or character interactions. Stereotype bias, a subcategory of unconscious bias, involves individuals’ evaluations being influenced by preconceived notions or stereotypes about their social group [177]. Similarly, some game-based assessments could counteract stereotype bias by providing a context-rich environment for candidates to display their competencies across various situations, thereby enabling a more comprehensive understanding of their capabilities. The game format redirects assessors’ attention towards candidates’ actions and decisions rather than stereotypes or superficial characteristics. Yet, game-based assessment implementation presents challenges, including the risk of inadvertently reinforcing stereotypes in game design [178] and the possibility that the diversity of game scenarios might not

encompass all skills and competencies requisite for specific job roles, potentially leading to incomplete candidate assessments.

Another threat an assessor may face in selection is the halo effect [179]. The halo effect refers to the tendency to attribute positive qualities to individuals based on their overall favourable impression, which can lead to biased assessments during the recruitment and selection process. Game-based assessment could be designed to help minimise the halo effect by providing objective, quantitative performance data that can be analysed and compared across candidates. This data-driven approach reduces the influence of subjective impressions, ensuring that candidates are evaluated based on their actual performance in the game rather than on superficial aspects that may create a halo effect. While such game-based assessment could minimise the halo effect by providing objective performance data, the interpretation of this data may still be subject to bias. When analysing the data, recruiters and hiring managers may be influenced by pre-existing beliefs or personal preferences, which could lead to biased decision-making.

Additionally, a crucial aspect of reducing bias in the recruitment and selection process is the promotion of objective decision-making among recruiters and hiring managers [180]. Some game-based assessments could support this by providing a standardised and structured assessment framework, which enables the comparison of candidates' performance based on predefined metrics and criteria. The data generated by the specified game-based assessments could be used to create evidence-based insights, which can inform the decision-making process and minimise the influence of subjective factors that could lead to biased judgements. Although game-based assessments could be designed to promote objective decision-making through standardised assessment frameworks, the reliance on quantitative metrics may not capture the full spectrum of candidates' skills and abilities, particularly in relation to soft skills and interpersonal competencies. As a result, the overemphasis on data-driven insights may inadvertently limit the comprehensive evaluation of candidates.

Finally, promoting diversity and inclusivity is essential to countering biases in the recruitment and selection process [181]. Specifically designed game-based assessment can foster an inclusive assessment environment by offering a variety of game genres and scenarios that cater to different interests, preferences, and cultural backgrounds. This approach ensures that candidates are provided with

equal opportunities to showcase their skills and abilities, irrespective of their differences, leading to a more equitable and unbiased selection process. While some game-based assessments aim to foster an inclusive assessment environment, the development of games that cater to a diverse range of interests and preferences can be resource-intensive and time-consuming [133]. Additionally, the accessibility of digital game-based assessment may be limited for candidates with disabilities or those who lack access to the necessary technology, which could inadvertently exclude certain groups from the recruitment and selection process.

In conclusion, specifically designed game-based assessments could offer a promising alternative to traditional recruitment and selection methods. Such assessments may also be designed to effectively reduce various biases and promote diversity while challenging the limitations in assessing soft skills and concerns about accessibility and inclusivity.

3.1.6 Conclusion

In conclusion, the use of specific game-based approaches in HR recruitment and selection has demonstrated promising potential in enhancing various aspects of the process. However, the academic literature has not yet fully captured the scope and implications of these rapidly evolving practices. Indeed, the academic literature mainly focuses on the advantages of game-based assessment for the recruitment and selection process. Consequently, no generic framework was proposed to create such game-based assessment means to our knowledge. Therefore, in the following section, we will discuss the strategy adopted in our research that involves the gamification of existing traditional methods, such as the situational judgement tests.

3.2 Game-based Soft Skill Assessment

Landers [182] and Hamari et al. [5] studies suggest that the gamification of existing methods rather than creating an entirely new serious game may be more accessible and more practical. Therefore, taking into account the latter suggestion and the contextual particularities presented in the previous paragraph, we have selected the following guidelines:

- Easily adaptable and scalable assessment means.
- User-friendly assessment methods that are simple to implement for diverse designers and stakeholders, ensuring seamless and cost-efficient integration and collaboration.
- Assessment means that promote soft skills and behavioural assessment through stealth assessment.

These guidelines led us to focus our research on the gamification of **low-fidelity simulations such as situational judgement tests**, initially studied by Georgiou et al. [49] based on simulated work-related scenarios in a game, who argue that these tests hold significant potential as valid game-based assessment tools.

3.2.1 Simulations for Recruitment and Selection

Simulations are components of candidate performance assessment and rely on the principle of **behavioural consistency**. This key concept contends that there is a correlation between a candidate's simulation performance and their real-world job performance [183, 184]. Interestingly, the application of this principle diverges between high-fidelity and **low-fidelity simulations**.

Low-fidelity simulations, such as situational judgement tests (SJTs), prioritise evaluating a candidate's procedural knowledge by understanding effective and ineffective behaviours within job-related scenarios. This type of knowledge is often deemed a reliable indicator of future job performance. Low-fidelity simulations assess the candidate's abstract understanding of appropriate actions across a range of professional situations without necessitating the direct performance of the tasks [183]. They present **hypothetical, job-related circumstances and require candidates to identify the most fitting response** or sequence responses based on their predicted effectiveness [31]. The scenarios can be conveyed through text or multimedia, with each scenario followed by multiple-choice questions.

For example, Oostrom et al. evaluate the soft skill '*Prudence*' from their SJT created in their study [185] and based on the HEXACO model:

Example of an SJT item

You are instructed to give a presentation to three board members. You have sufficient time to prepare the presentation. What would you do?

- A) I would start preparing the presentation an hour in advance because I perform best under pressure. (1)
- B) I would prepare the presentation a few days in advance and then check and practice the presentation at the last minute. (2)
- C) I would finish the presentation way in advance and carefully check the presentation. I would practice the presentation at the last minute. (3).
- D) I would finish the presentation way in advance. I would carefully check the presentation a few times myself and then have a colleague look at it as well. Way in advance, I would run through the presentation a few times so I know it by heart. (4)

In this example, the preferred response would typically be option *D*, as the researchers noted by ranking the item answers from *1* to *4*, with *1* as the least preferred answer and *4* as the most preferred answer.

While not designed as direct measures of personality, research implies that SJTs can shed light on various personality traits, as argued by Cabrera et al. [186], who highlight connections between SJTs and traits like conscientiousness, emotional stability, and extraversion.

It is crucial to differentiate SJTs from self-report personality tests such as the Myers-Briggs Type Indicator (MBTI), Big Five, or HEXACO surveys. Although they both involve personal insight, self-report tests prompt individuals **to reflect on their behaviours, feelings, and attitudes and provide self-descriptions**. In contrast, SJTs **necessitate judgement and decision-making skills in response to specific scenarios**, making them less subjective and more reflective of a candidate's likely real-world behaviour.

Conversely, high-fidelity simulations, like those employed in assessment centres, **prioritise the evaluation of a direct candidate's behaviours in a situation**. These include in-basket exercises, where participants respond to typical job-related

tasks, role-playing scenarios that evaluate interpersonal skills, group exercises that assess teamwork and leadership capabilities, and business games that test strategic decision-making and financial management [187]. Additionally, virtual reality simulations are used for roles requiring specific technical skills, providing a realistic work environment [188]. The choice of simulation offers candidates a platform to display their procedural knowledge and skill set through tangible behaviour [183]. They involve more **realistic, immersive, task-specific scenarios that closely mirror the job environment**. The behaviours observed during these simulations are then extrapolated to predict future on-the-job behaviour [189]. Despite the comparable psychological fidelity between the Situational Judgement Tests (SJT) and assessment centers utilised in this research, Blair et al. highlight in their study that the assessment centers may prove to be a more **reliable and accurate predictor of outcomes over an extended period** [190].

In conclusion, low-fidelity simulations emphasise the assessment of a **candidate's understanding of effective job-related behaviours**. In contrast, high-fidelity simulations permit **the direct observation of these behaviours in lifelike settings**. Both simulation types aim to **predict future job performance based on behavioural consistency** but use different evaluative approaches to achieve this goal with high-fidelity simulations that tend to be more valid than low-fidelity simulations over time. Yet, high-fidelity simulation may not follow the guidelines expressed in the premises of the section as such assessment means are long, tedious and not scalable by their nature.

Given the strengths and weaknesses of both high-fidelity and low-fidelity simulations, our research will focus on the application of **gamified situational judgement tests (GSJT)**. This approach may leverage the advantages of both simulation types, aligning with our guidelines. Gamified SJTs may provide an engaging, scenario-based evaluation platform similar to high-fidelity simulations, allowing for **the observation of behaviours in a simulated environment**. At the same time, they may maintain **the convenience, scalability, and focus on procedural knowledge inherent in low-fidelity simulations**. Therefore, a GSJT may be categorised as a *mid-fidelity simulation*.

3.2.2 Gamified Situational Judgement Tests

Gamified situational judgement tests (GSJT) essentially are SJTs that employ game-like features and present candidates with authentic, work-related scenarios, requiring them to select the most appropriate response from multiple options (see Figure 3.7). By incorporating game elements such as **narrative, role-playing, and challenges**, these tests strive to provide a **more engaging and enjoyable experience** for candidates while **effectively evaluating their decision-making**, as well as **job-relevant soft and interpersonal skills** [49].

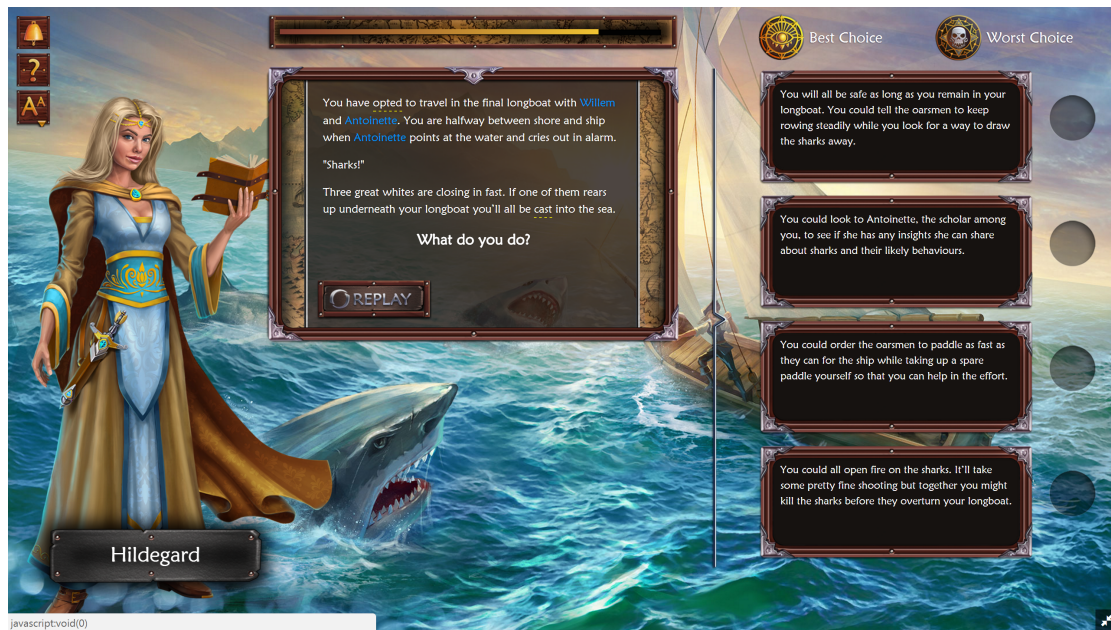


FIGURE 3.7: Example of a Gamified Situational Judgement Test for soft skill recognition by Owiwi ¹

Moreover, the gamified situational judgement tests align with the guidelines chosen for our research domain, as they offer **easy adoption and intuitive gameplay through realistic work-related situations**. They also provide **adaptability and scalability** through modifiable external resources, such as scenarios and graphic assets. Additionally, the tests implement **stealth assessment** with a game dynamic that stimulates participant motivation based on **narrative and interactive activities**, which, in turn, promotes **test anxiety reduction** [49].

While gamified situational judgement tests (GSJT) are progressively gaining traction in psychometric assessment in industries, it is important to underscore that

¹OWIWI is a Greek company mainly activated in serious games' software production for soft skill recognition.

the body of research surrounding this innovative approach is still in **its formative stages**. As the concept is relatively new, empirical studies devoted explicitly to understanding, designing, applying, and evaluating GSJT are relatively limited compared to the extensive body of literature on traditional SJTs [191].

However, through their literature review, Aouam et al. [191] examined GSJT research, along with other game-based assessment approaches research for personnel and recruitment selection. The outcomes of this exploration illuminated four primary themes prevalent within the body of knowledge related to gamification in the context of personnel selection.

- The first theme revolves around **candidate reactions and perceptions towards gamification tools and serious games**. It was found that gender significantly influences candidates' perceptions, with men generally reacting more positively than women. Moreover, these tools were shown to have a high-reliability level and could assess various soft skills, such as diligence, self-control, and task orientation. Additionally, candidates' attitudes towards gamification correlate positively with their satisfaction levels, intentions to recommend, and perceived organisational attractiveness.
- The second theme explores **applicant responses to gamified selection tests**. Empirical studies showed no significant influence of gamified situational judgement tests on applicants' perceptions of fairness and organisational attractiveness. However, using these tests contributes to a higher favorability rating than traditional selection tests, leading candidates to perceive the organisation as innovative.
- Third theme focuses on **candidate performance within a gamified selection process**. Research suggests that gamified selection tools can predict job performance by measuring soft skills such as adaptability, flexibility, and decision-making. Moreover, gender was found to influence candidates' performance in a gamified selection process.
- The final theme discusses **the benefits and limitations of gamification**. Using gamification in the personnel selection process decreased candidate stress and anxiety while increasing motivation and involvement. Moreover, gamified selection tools can detect specific soft skills often overlooked by traditional selection tools. However, applying these tools requires considerable

financial resources and implementing a robust design to ensure tool validity and reliability.

Despite the preliminary positive results from the literature review, the relative scarcity of research in the game-based assessment research domain for recruitment and selection presents a dual challenge. Firstly, it creates **a knowledge gap in understanding the full range of benefits, potential pitfalls, and optimal design strategies associated with GSJT**. Secondly, it limits the capacity to **draw definitive conclusions about their validity and reliability in evaluating soft skills and behavioural traits** [192]. Finally, **no framework to create a game-based assessment approach is proposed to generate a gamified situational judgement test**.

The need for more rigorous and comprehensive research on GSJT is thus evident. Expanding the research domain will provide valuable insights into **the potential of GSJT in different contexts and populations and contribute to refining their design and implementation for enhanced effectiveness**. This would consequently **enrich the methodological toolkit of psychometric assessment** and deepen the understanding of **the utility of gamification in the realm of psychological testing** [133, 193].

Given the current landscape of gamified situational judgement test (GSJT) research, with its manifold yet largely unexplored opportunities and challenges, a compelling need exists to expand the research terrain. This can be achieved by carefully exploring various facets of GSJTs, particularly their creation and design processes. In this light, a contribution of the thesis focuses specifically on these aspects of GSJTs design and production. In so doing, we aim to add to the existing knowledge base and potentially catalyse further research activity in this emerging field. **Our goal is to facilitate a more seamless integration of gamification principles into psychometric assessment practices, thereby advancing the science and utility of psychological testing through GSJTs based on interactive narrative game genre.**

3.2.3 Interactive Narrative Game and Gamified Situational Judgement Test

Interactive Narrative Games (ING) encompass a distinct genre of video games that prioritise **the narrative structure and players' agency**. In these games, players take on the role of a protagonist, engaging with other characters and exploring various environments within the game world. Players can influence the narrative's progression and direction through **dialogue options, actions, and other interactions, including different endings, character outcomes, and story consequences**. ING can manifest in several forms, such as **role-playing games, point-and-click adventure games, or visual novels**. These games are known for their **intricate characters, profound storylines, and captivating environments**, enabling players to become fully **immersed in the game world**.

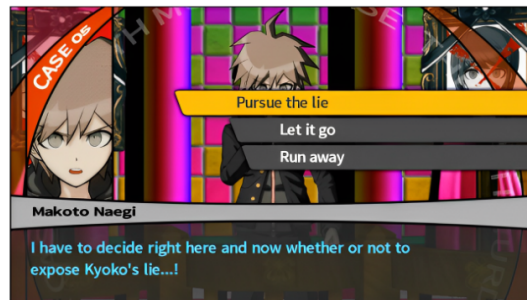
Well-known examples of ING include the Danganronpa series, Life is Strange, and The Walking Dead series (see Fig 3.8). These games have garnered acclaim for their inventive storytelling and ability to engage players in emotionally impactful narratives.



Life is Strange



The Walking Dead



Danganronpa

FIGURE 3.8: Example of Interactive Narrative Games: Life is Strange, The Walking Dead, Danganronpa)

There are noteworthy similarities between ING and traditional Situational Judgement Tests (SJT). Both activities heavily rely on **decision-making as a core element**. Furthermore, storytelling is a crucial aspect that captivates players or test-takers in ING and SJT. INGs are renowned for their enveloping narratives, emphasising character development and granting players a sense of agency within the game world. On the other hand, SJT employs realistic situations to provide a narrative context that engages test-takers and prompts them to consider various factors affecting decision-making in real-life circumstances.

Lastly, ING and SJT provide highly engaging and immersive experiences. The ability of these activities to transport players or test-takers into alternate realities or situations is a significant factor contributing to their allure. Understanding the parallels between ING and SJT can inform the development of gamified assessment tools, such as Gamified Situational Judgement Tests (GSJT).

Consequently, our work is centred around **the development of game-based assessment, drawing from interactive narrative games, aimed at evaluating a wide range of diverse candidates**.

Having defined ING and GSJT with their significance for recruitment and selection processes, we will now explore the challenges of devising such assessment tools.

3.3 Toward the Design and Conception of GSJT

As previously elucidated, the concept of Gamified Situational Judgement Tests (GSJT) constitutes **a relatively nascent development in academic literature**. As such, the present section aims to interrogate the breadth of extant research on this subject. We will examine **the complexities of designing a Situational Judgement Test (SJT)**, with a particular emphasis on **the underlying scoring systems**. This aspect of our study is particularly pertinent, as it has the potential to impact the validity and efficacy of GSJT as a research instrument.

Simultaneously, our investigation will explore and elucidate the principles guiding the creation of ING as the chosen medium in our thesis for using the GSJT approach in soft skills and behavioural recognition.

By adopting this dual-pronged approach to the literature review, we hope to triangulate data from disparate sources, thus providing a holistic perspective. This comprehensive exploration of the intricacies of SJTs and the design and conception process of ING will serve as a foundation upon which we intend to propose an original solution, thereby addressing the research questions that motivate this study.

3.3.1 Stakeholders involved in Situational Judgement Test

We highlighted earlier that situational judgement tests (SJTs) are increasingly employed in recruitment and personnel selection due to their effectiveness in predicting job performance and their capacity to assess soft skills, such as communication, problem-solving, and interpersonal abilities [194].

The process of SJT development requires a team of specialists with complementary skills and knowledge. Primary contributors often include **industrial-organisational psychologists, subject matter experts (SMEs), and human resource (HR) professionals**.

- **Industrial-organisational psychologists** bring their understanding of human behaviour within workplace contexts to the design and development of SJTs. Their expertise often guides the item generation and validation process, ensuring that the tests are reliable and effective measures of the attributes they intend to assess [195].
- **SMEs**, with their in-depth understanding of the roles for which the SJTs are being developed, contribute significantly to constructing scenarios and response options that genuinely reflect the situations that could potentially occur in the respective jobs. Their involvement thus ensures the content validity of the SJT [196].
- **HR professional** 's role is to ensure alignment of the SJTs with the organisation's overall recruitment and selection objectives. They provide valuable insights about the practical considerations of administering the SJT, such as test duration, format, and available resources for its implementation [49].

3.3.2 Design and Conception of Situational Judgement Tests

Working together, the stakeholders generally follow a structured, multi-stage process (see Figure 3.9), beginning with a **comprehensive job analysis** to generate an SJT [197].

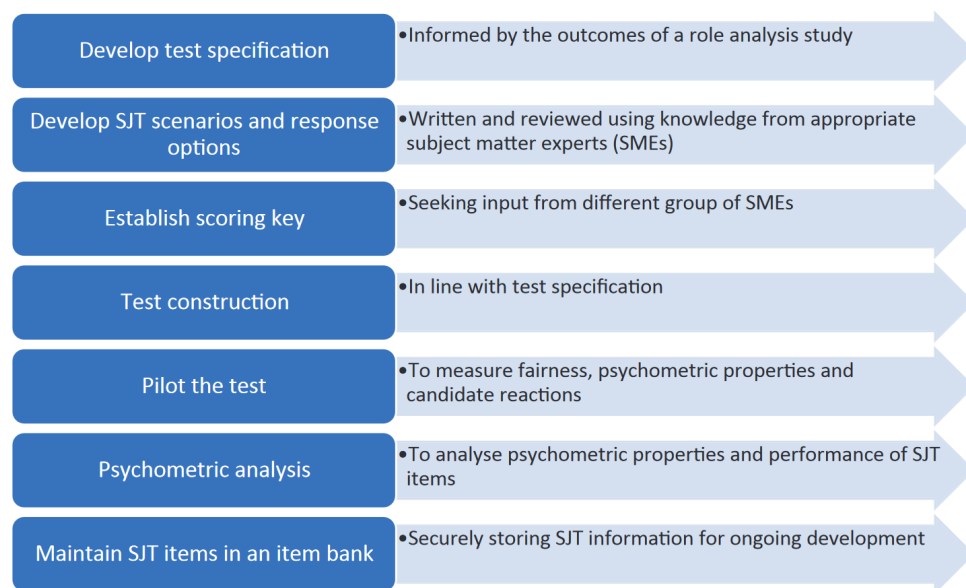


FIGURE 3.9: Situational Judgement Test Design and Production Process (From Patterson and Driver, 2018, p. 82)

This analysis identifies key competencies or constructs that the SJT will assess, utilising observations, interviews, and surveys to gather data about the job's tasks and the knowledge, skills, and abilities required for effective performance.

Following job analysis, **item generation is undertaken**. This involves creating a series of job-related situations and corresponding potential responses. These situations aim to reflect the intricacies and challenges inherent in the job role, and the responses typically represent varying degrees of effectiveness.

Once items are generated, **scoring the items' responses** is carried out. Several methods exist to score the item's responses, such as consensus, expert-based, and rational scoring. Following our objective to assist psychometric experts in developing a relevant in-game soft skill measurement system, we will provide further and extensive details about the different methods in the chapter 6, section 6.1.

Then, the SJT undergoes **test construction** in line with test specifications and pilot testing with a representative sample of the target population. Statistical

analyses assess item functioning and test reliability. Based on the results, necessary revisions are made to the items.

Finally, **psychometric analyses** are performed to check the reliability and validity of the items before being stored in an item bank.

In conclusion, the design and conception of SJTs necessitate an amalgamation of expertise from industrial-organisational psychology, HR, and subject matter expertise underpinned by a rigorous and systematic methodology. When properly developed, SJTs may offer a robust, valid, and effective personnel selection and recruitment tool.

3.3.3 Design and Conception of Interactive Narrative Games

Interactive narrative games signify a distinct sphere within the larger framework of digital entertainment. In these games, the two foundational elements, player agency and interactive storytelling intertwine seamlessly to engender robust, immersive experiences [198]. Unlike conventional games, where the story often forms a backdrop, interactive narrative games place narrative as a central pillar of the gaming experience. The design of these games is thus an exercise in achieving equilibrium between game mechanics and the unfolding of the narrative, striving to deliver an experience that rewards players not just through gameplay but also through **narrative progression** [199].

Designing interactive narrative games involves understanding and implementing principles from disparate disciplines such as game design, narrative theory, and psychology. A fine balance must be achieved where the mechanics of the game and the narrative structure weave together into a compelling whole that draws the player in and holds their attention. This balance allows players to feel that their participation in the game universe is meaningful and consequential, further enriching the overall gameplay experience.

Narrative structures within interactive games can be categorised broadly into four main types: **linear, branching, modular and emergent** (see Figure 3.10):

- **Linear narratives** echo traditional literature and film, following a fixed storyline with predetermined events [200]. This structure ensures a coherent,

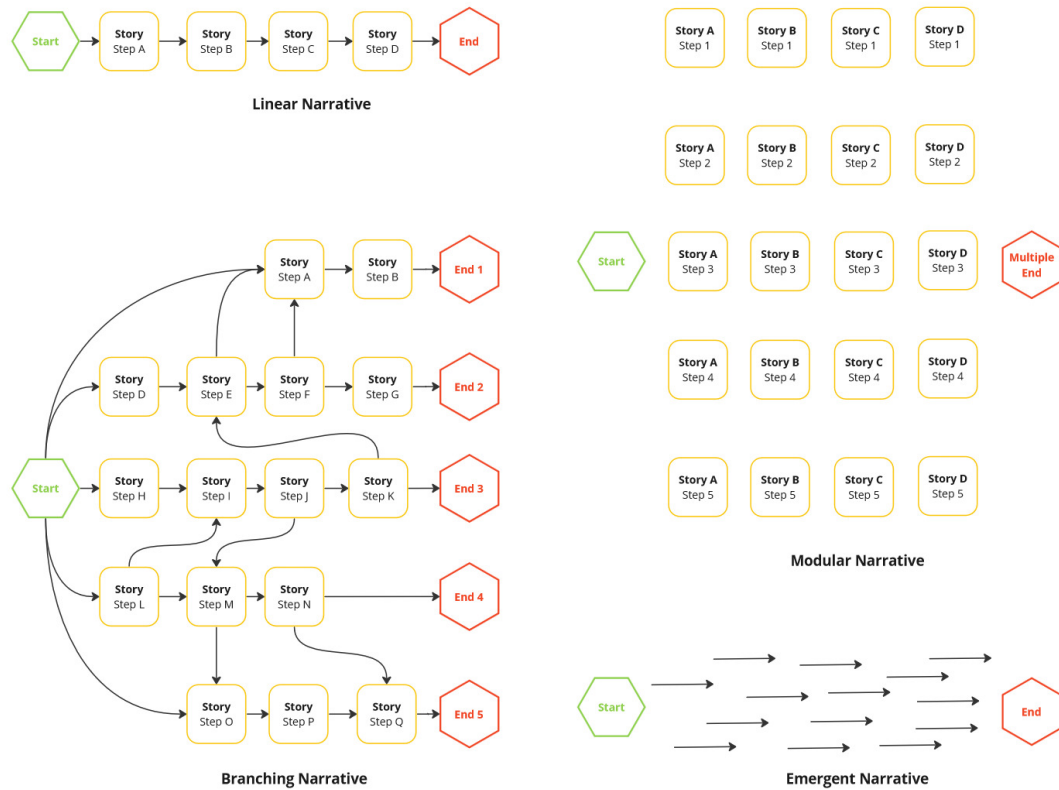


FIGURE 3.10: Narrative Structure Categories

author-driven narrative but potentially limits player agency and the ability for unique player experiences.

- On the other hand, **branching narratives** provide players with the opportunity to directly influence the progression of the story. Player decisions can lead to different outcomes and divergent paths, thus offering a higher level of agency [201]. However, the design of such games can quickly become intricate, as each decision can exponentially increase the number of potential narratives, a phenomenon referred to as a 'combinatorial explosion.'
- A **modular narrative** structure refers to a storytelling structure organised in self-contained, interchangeable segments or modules. Unlike linear or branching narratives that follow a continuous, sequential path, modular narratives can be reassembled in different orders, offering multiple paths and outcomes for the story. Each module serves as an autonomous unit with its own arc, contributing to the overall narrative while simultaneously being capable of standing independently.

- **Emergent narratives** offer a different approach altogether, which has been gaining traction in recent years. Instead of scripting the narrative in advance, emergent narratives rely on systemic interactions within the game world to generate storylines. These narratives can be highly unique and unpredictable, providing the player with a sense of trustworthy agency. However, this structure's inherent unpredictability may make maintaining narrative coherence and consistency challenging. Therefore, such an approach may be applied through the use of a drama manager, which plays a central role in managing and adjusting the narrative arc of a story in response to the actions and decisions of the player, creating a more dynamic, immersive and personalised interactive experience[202, 203].

3.3.3.1 Drama Manager in Interactive Narrative Games

While linear and branching narratives are **pre-scripted structure** that require no real-time and in-game management, the modular and emergent narrative is **dependent on the player agency**; therefore, a drama manager is needed to control them. Though not new, the concept of a ***Drama Manager*** has evolved significantly over the years. The earliest notion of a *Drama Manager* can be traced back to work by Bates [203] and Laurel [202], who envisaged the role of a computer as a director of a drama, suggesting that a computer could **manipulate the dramatic unfolding of an interactive experience** much like a theatre director. This idea provided a foundation for later work on **interactive storytelling** and paved the way for the development of the Drama Manager (DM).

Drama management in the field of interactive narrative games refers to an artificially intelligent system tasked with the responsibility of dynamically adjusting the progression of the game's narrative in response to the actions of a player [204]. A drama manager's role can be akin to that of a digital 'dungeon master'², guiding the player through the narrative, adapting it on-the-fly, ensuring the overall cohesiveness and engagement of the story [205].

The primary objective of a drama manager is **to ensure that the story remains meaningful and coherent irrespective of player actions**, thereby enhancing

²In the Dungeons & Dragons (D&D) role-playing game, the Dungeon Master (DM) is the game organiser and participant in charge of creating the details and challenges of a given adventure while maintaining a realistic continuity of events. - https://en.wikipedia.org/wiki/Dungeon_Master

player engagement. The DM also facilitates player agency by allowing players to influence the story through their actions.

There are several techniques that drama managers may utilise to shape the narrative. These may include but are not limited to variable control, plot graph manipulation, character behaviour modification, and world-state transformation [206]. For example, in *Facade*, a popular interactive drama game in the domain, the drama manager assesses the player’s dialogue inputs and manipulates the characters’ responses to guide the narrative [204].

As technology and AI methodologies evolve, the complexity and efficacy of drama managers have also grown. Recent advancements include machine learning techniques for more nuanced player modelling and the employment of procedural content generation for creating diverse narrative elements [207]. Such advancement in the domain may enable the development of more sophisticated player models that capture the player’s emotional state and cognitive traits and could enable the DM to craft narratives that resonate more deeply with the player, paving the way for genuinely personalised interactive experiences.

However, the challenge of striking the perfect balance between player agency and narrative coherence remains a key issue in the design of drama managers, especially when **several non-playable characters** are involved in the story.

3.3.3.2 Non-Playable Characters in Interactive Narrative Games

Non-playable characters (NPCs) in interactive narrative games serve a pivotal role **in augmenting the depth and realism of the game world**. They influence player immersion and engagement, assist in advancing the storyline, and contribute to the complexity of the game’s decision-making and strategic elements [208].

Indeed, NPCs significantly enhance the player’s immersion and engagement by contributing to the world-building aspect of the game. By exhibiting behaviours consistent with the game’s setting and narrative, NPCs provide a sense of authenticity and depth to the virtual world. This can lead to a more immersive and compelling experience for players as they interact with characters who react realistically within the context of the game world [209].

Additionally, NPCs often act as **drivers of the game's narrative**, delivering crucial information, quests, or challenges to the player. They serve as a conduit for the game's story, providing exposition, context, and motivation for the player's actions. Through interactions with NPCs, players can uncover layers of the game's narrative, enhancing their understanding of the game's plot and its underlying themes [210].

Finally, NPCs can also contribute to **the strategic and decision-making aspects** of interactive narrative games. Complex NPC behaviour can provide a challenge and stimulate strategic thinking, as players must consider how their actions might influence the NPC's reactions. In games where the narrative branches based on player choices, NPCs' responses to player decisions can lead to different story outcomes, **enhancing the game's replayability** [200].

Finally, NPCs can provide **a nuanced moral and ethical dimension to the game**. Through their interactions with the player and reactions to the player's decisions, NPCs can highlight the moral and ethical implications of the player's actions, thus enriching the narrative and emotional depth of the game [211].

In summary, the importance of NPCs in interactive narrative games extends beyond merely populating the game world. They contribute **to immersion and engagement, drive the game's narrative, facilitate strategic decision-making, can imbue the game with moral and ethical complexity**. Therefore, effective design and integration of NPCs are crucial for creating rich, engaging, and dynamic interactive narrative games. Furthermore, these NPCs possess the potential to significantly influence player agency, further emphasising their significance in the overall gaming experience.

3.3.3.3 Player Agency, Assessment System and Interactive Storytelling

Player agency, defined as **the player's ability to make decisions** that significantly influence the game world, is a critical element in the design of interactive narrative games. Ensuring player agency requires striking a balance between offering players meaningful choices and maintaining the integrity of the narrative. The narrative might become disjointed or incoherent if the game offers too many choices. Conversely, too few choices can leave players feeling like passive observers rather than active participants. Facade (see Figure 3.11), a game developed by

Mateas and Stern [204], is an example of successfully striking this balance. In *Facade*, an AI-driven drama manager dynamically adjusts the narrative progression based on the player's actions, balancing player agency and narrative coherence.



FIGURE 3.11: *Facade* game by Mateas and Stern

The concept of player agency in interactive narrative games, as *Facade* exemplifies, provides a range of unique advantages for recognising soft skills in the recruitment and selection process.

One critical aspect of player agency is the power of players to make meaningful choices, significantly influencing the game world. These decisions can offer substantial insights into an applicant's decision-making abilities. This includes their capacity for strategic thinking, balancing multiple options, and considering the potential consequences of their actions. Such observations may not be readily apparent in more conventional recruitment contexts.

Furthermore, the balance *Facade* strikes between player agency and narrative coherence **promotes active participation over passive observation**. In an environment where the narrative adjusts dynamically based on the player's actions, the candidate's adaptability can be observed, indicating their potential response to change, strategy adjustments in response to new information, and management

of unexpected situations. This necessitates an intelligent, embedded assessment system to interpret player behaviour within the gameplay experience.

In conclusion, the concept of player agency in interactive narrative games can be a dynamic, engaging, and insightful tool for recognising and assessing soft skills during the recruitment and selection process. Key attributes and skills can be revealed by providing an environment where candidates' decisions and interactions have a real impact. The balance between player agency, assessment system and narrative coherence may be essential for the efficacy of an innovative approach to fair and coherent recruitment.

3.3.4 Stakeholders involved in Interactive Narrative Games

The conception phase of interactive narrative games marks the beginning of the game development cycle. It involves generating and refining the game's initial idea, defining its primary components, and planning its eventual execution [212]. This phase is crucial, as it sets the foundational framework upon which the game will be built, and it involves various stakeholders, including game designers, narrative designers, artists, and programmers:

- **Game Designers:** Game designers are responsible for designing the game mechanics and rules that will drive the gameplay. They must consider how these mechanics interact with the narrative to create a satisfying and coherent player experience.
- **Narrative Designers and writers:** Narrative designers and writers focus on crafting the game's story and how it unfolds based on player decisions. They often work closely with game designers to ensure the narrative and gameplay are tightly integrated.
- **Artists and Creative population:** Artists and creative population contribute to the game's aesthetic, creating the visual elements that bring the game world to life. This includes character design, environment design, and user interface design. They must cooperate with the narrative and game designers to ensure a cohesive vision.
- **Programmers:** Programmers are responsible for coding the game mechanics, physics, and artificial intelligence, along with overseeing the monitoring

and analysis of player data within the game. They also develop tools to assist in the design and testing process. Their role in the conception phase often involves planning the technical architecture and establishing the feasibility of proposed game mechanics.

Each company may propose its own methodology and framework depending on its resources and working routines. Yet, to help these stakeholders, several methodologies may guide the conception phase of game development. These are not strict, inflexible rules but frameworks to aid in structuring and streamlining the conception process.

- **Iterative Design:** This is a cyclic method of designing, prototyping, testing, and refining a game. This methodology is particularly suited to game design, as it allows for continual improvement and refinement based on feedback [212].
- **Agile Development:** Agile development methodologies, such as Scrum, emphasise flexibility, collaboration, and incremental development. These methodologies can be advantageous for interactive narrative game development due to the inherent uncertainty and complexity of designing complex narrative structures and gameplay mechanics [213].
- **User-Centered Design (UCD):** UCD focuses on designing games considering the end user's needs and preferences. This approach can be beneficial in interactive narrative game design, as it can help to ensure that the player's agency and the game's interactivity are well-integrated into the narrative structure [214].

In conclusion, the design and conception phases of interactive narrative games involve **a complex, collaborative effort from various stakeholders**, each bringing their expertise to the table. The use of **structured methodologies can help guide this process**, aiding in coordinating the different elements and ensuring that the game's narrative and interactive elements are well integrated.

3.4 Conclusion

In the first part of this chapter, we firstly explored **the adoption of game-based approaches in recruitment and selection processes** engenders multifaceted advantages that enhance both the quality and efficiency of the process. Firstly, **an enhanced candidate experience** is offered, typically fostering greater engagement and motivation among applicants. This emanates from the immersive and interactive nature of the games, which diverges from traditional, often monotonous methods.

Secondly, **the candidate pool is broadened** as game-based methodologies, being intrinsically attractive and novel, draw more diverse and numerous applicants. This heightened appeal can encourage those who may have initially been reluctant to engage in a formal recruitment process.

The third advantage pertains to eliciting **improved applicant reactions to the selection process**. Game-based methodologies are often perceived as more fair and enjoyable, thus engendering positive attitudes and fostering improved organisational perceptions.

Fourthly, the provision of **behavioural measures are facilitated**, as game-based selection methods provide an opportunity for applicants to display real-world skills and abilities, rather than simply relying on traditional methods such as cognitive testing or interviews, which may not always capture the full breadth of a candidate's potential.

Finally, game-based approaches may significantly mitigate the presence of assessment biases. This can occur by establishing standardised, objective scoring metrics less susceptible to subjectivity and bias than traditional recruitment methodologies. Moreover, because these games are typically grounded in scientific research and psychometric principles, they may also provide a more accurate representation of a candidate's ability, thus enhancing the fairness and validity of the selection process. In essence, game-based recruitment and selection offer a modern, robust, and innovative solution to the challenges of traditional hiring practices, contributing to a more efficient, fair, and engaging recruitment experience.

In the second part, we revealed **the nascent gamified situational judgement test approaches** used by the industries that want to propose the advantages of

game-based approaches mixed with **an easily adaptable, easy-friendly assessment means**. Indeed, we explored that a GSJT is a modern assessment tool that combines the principles of **gamification with the traditional structure of SJTs** to evaluate some individual's soft skills and behavioural tendencies. Gamification entails incorporating game-like elements into non-game contexts to foster engagement and motivation.

In the context of recruitment, gamified SJTs present hypothetical scenarios relevant to the job role, and candidates are asked to respond or make decisions based on the situation presented. The responses provided offer insights into the candidates' soft skills. These tests are constructed **to mimic real-life situations, making them more immersive and allowing for a more authentic and holistic assessment of candidates' abilities**. The gamified element enhances candidate engagement and can result in a richer, more comprehensive data set for employers to evaluate. Thus, GSJTs may provide a more precise and objective measure of a candidate's potential for success in a given role, especially regarding their soft skills and behavioural attributes.

Hence, we describe in the third part **the design and the conception process of both components constituting a gamified situational judgement test**. On one hand, we have the **SJT**, and on the other hand the **interactive narrative game genre**. In addition, we highlighted the design and production process of a GSJT. We highlighted three important concepts in the design and production of a GSJT.

Firstly, the drama manager suggests that the role of a system as a drama director could manipulate the dramatic unfolding of an interactive experience, much like a theatre director. The primary objective of a drama manager is to ensure that the story remains meaningful and coherent irrespective of player actions, thereby enhancing player engagement. The DM also facilitates player agency by allowing players to influence the story through their actions.

Second, we highlighted the importance of NPCs in ING and GSJT. NPCs in interactive narrative games hold importance that transcends the mere population of the game world. They facilitate immersion, drive the narrative, enhance strategic decision-making, and can introduce moral and ethical complexity needed in a GSJT. Consequently, their effective design and integration are essential for crafting engaging and dynamic interactive narrative games. Furthermore, their potential

to profoundly affect player agency further underscores their importance in the overall gaming and assessment experience.

Finally, the coherence between player agency and the goal of GSJT to assess soft skills has also been explored. We highlighted that player agency in interactive narrative games, defined as a player's power to make decisions that impact the game world, is vital for game design as well as the measure of soft skill. This agency also requires a careful balance between giving players meaningful choices and maintaining the narrative's integrity. A carefully designed GSJT can create an environment where candidates' decisions have a real impact, revealing key attributes and skills while maintaining a balance between player agency, assessment system, and narrative coherence.

However, to our knowledge, no framework has been proposed to generate a soft skill assessment for GSJT. Hence, given the nascent nature of domain research, our thesis aims to contribute to achieving our objectives by assisting designers in developing a GSJT specifically tailored for recruitment and selection processes.

Within the forthcoming chapters, we introduce frameworks and models crafted to navigate the research context. This context is centred on **the facilitation of a multidisciplinary team in conceiving a gamified situational judgement test (GSJT)** designed to recognise behavioural profiles and soft skills.

Furthermore, our investigation delves deep into the previously outlined three pivotal dimensions of GSJTs. These dimensions encompass **the refinement of the scoring system, ensuring both precision and equity, the artful construction of lifelike non-player characters (NPCs) to infuse authenticity into the interactive realm, and the orchestration of a dynamic, captivating narrative structure.** These elements converge to yield an unparalleled and immersive encounter that stands as genuinely distinctive.

Chapter Overview

In this chapter, we observed that despite businesses adopting game-based approaches for potential benefits, the academic literature on the topic is still in its nascent stages and does not fully address our research objectives as outlined in Chapter 1. Thus, we explored the subject of **gamified situational judgement tests** in this chapter, highlighting the necessity for substantial collaboration within a **multidisciplinary team**. Our focus was on exploring research areas that could assist in the **development of a game-based assessment for behavioural profiling**, aligning with the primary objective of our thesis to **aid GSJT designers**.

Research Objectives:

- **Objective A: Develop a game-based approach for immersive behavioural assessment**
- **Objective B: Propose innovative strategies to streamline the gamification process.**

Research Areas:

- **Gamified Situational Judgement Tests approach for behavioural recognition**
- **Frameworks to aid a multidisciplinary team in generating a GSJT**
- **Drama Manager for narrative coherence and dynamic tension between player agency and authorial control in GSJT**
- **Computational Psychometric Scoring Approaches for GSJT scoring processes**
- **Machine and deep learning models to create realistic non-playable characters in GSJT**

Chapter 4

Proposing a Meta-Framework and a Preliminary Framework to Generate a Gamified Situational Judgement Test

The previous chapter mentioned that gamified situational judgement test design is quite **new in academia**. Despite its clear potential, there is a noticeable lack of extensive research and development in this area. As a result, there is a knowledge gap that could benefit from a more structured approach to handle the complexities of creating such an assessment tool.

To bridge this gap, we chose to conceptualise the generation of GSJT through a **structured framework architecture and a methodology capable of delineating the nuanced process of transmuting a traditional SJT into an engaging game**. The goal is to propose overcoming the limitations of an isolated viewpoint and adopting a more holistic method. This approach would simultaneously recognise and tackle the complex challenges of converting SJTs into serious games.

With this backdrop, we propose a **preliminary meta-framework and a framework to generate a gamified situational judgement test**. While the meta-framework aims to give GSJT designers a **global overview and key guidelines to create a GSJT** or the possibility to create their framework, the framework proposal, based on the meta-framework, aims **to equip designers with a holistic,**

robust, and versatile tool, encapsulating a detailed procedural blueprint for designing and producing a GSJT. The proposed structure represents a generalised approach to gamification.

This framework not only embraces an integrative approach that considers the larger picture of SJT gamification, thereby setting a new precedent in the academic discourse surrounding GSJT design and production but also offers a representation of the actors and the roles involved in the creation of a GSJT previously presented in section 3.3.

4.1 Meta-framework for designing and creating Gamified Situational Judgement Test

In this section, we will first provide an overview of the meta-framework and then elaborate on our preliminary proposal for a comprehensive methodology of collaborative GSJT.

We will identify the roles of various actors involved in the GSJT creation process and outline each phase of the methodology, highlighting the specific contributions of each role. The elements presented in this section draw upon our literature research, the field experiment, and an analysis of practices from a video game company.

Hence, we outline in the following a GSJT generation process with different stages, described in Georgiou et al.'s study [49]:

- **Identification of Core Competencies and Skills:** The researchers first identified the most common skills and competencies organisations look for when hiring, especially for entry-level and graduate positions. These included adaptability, flexibility, learning agility, knowledge breadth, multicultural perspective, decision-making, resilience, and ability to work under pressure.
- **Selection of Skills for Assessment:** They then selected four increasingly relevant skills in contemporary work environments: resilience, adaptability, flexibility, and decision-making. These skills were chosen to form the dimensions of the Situational Judgement Test (SJT) and the gamified assessment.

- **Decision to Use Gamified Assessment:** The researchers decided to use a gamified assessment due to the difficulty of transferring and assessing soft skills compared to hard skills. They also used references to previous studies to support the suitability of gamified assessments for these kinds of skills.
- **Choice of Gamification Type:** The authors then chose the type of gamification to use, drawing from a taxonomy of gamification elements used in educational contexts. They followed certain gamification design principles: engagement, feedback, progress, freedom of choice, and storytelling.
- **Development of Gamified SJT Assessment:** The authors developed the gamified SJT assessment using game mechanics that served the above-mentioned principles.
 - For instance, in their research, Georgiou et al. [49] propose to the test-takers to select a hero/avatar and provide each character with a back-story. The story follows the heroes on a journey through four islands, one for each soft skill assessed. Visual and voice-overs are used for storytelling to increase engagement. The researcher also displays a visual progression bar that shows the test-takers' progress in the assessment, and a world map shows the islands they have traversed. Moreover, test-takers rewards are intrinsic (completing the missions) and extrinsic (receiving feedback on their competencies after the game). Finally, the test-takers can choose their avatar, skip the narrative, and leave the game anytime, continuing later from where they left off.
- **Balancing Assessment and Game Mechanics:** The authors balanced the assessment and game mechanics to make the experience fun and engaging without alienating non-gamers and to ensure fair and equal opportunities for all test-takers.

Our following proposal of meta-framework is based on the research of Georgiou et al. [49] and relies on **three dimensions: Behavioural Psychology, Situational Judgement Tests, and Gamification** (see Figure 4.1).

- **The Behavioural Psychology** dimension serves as the foundational layer for understanding individual responses in specific situations. Rooted in various psychological theories, this area is key for assessing personality traits

and behavioural patterns. By incorporating elements from human resources, educational psychology, and data analytics, a nuanced understanding of behaviour within a professional or academic setting is achieved.

- **The SJTs dimension** is the core of the assessment methodology. Derived from psychometrics and cognitive psychology, this area focuses on crafting realistic scenarios that require judgement, decision-making and soft skills engagement. The domain encompasses job analysis and assessment technology principles to ensure the tests are valid and easily administered.
- **The Gamification dimension** adds a layer of engagement to traditional SJTs. Borrowing concepts from game design and motivational psychology, this area seeks to enhance user interaction and data capture. It is not merely an overlay but an integral component enabling stealth assessment for behavioural recognition.

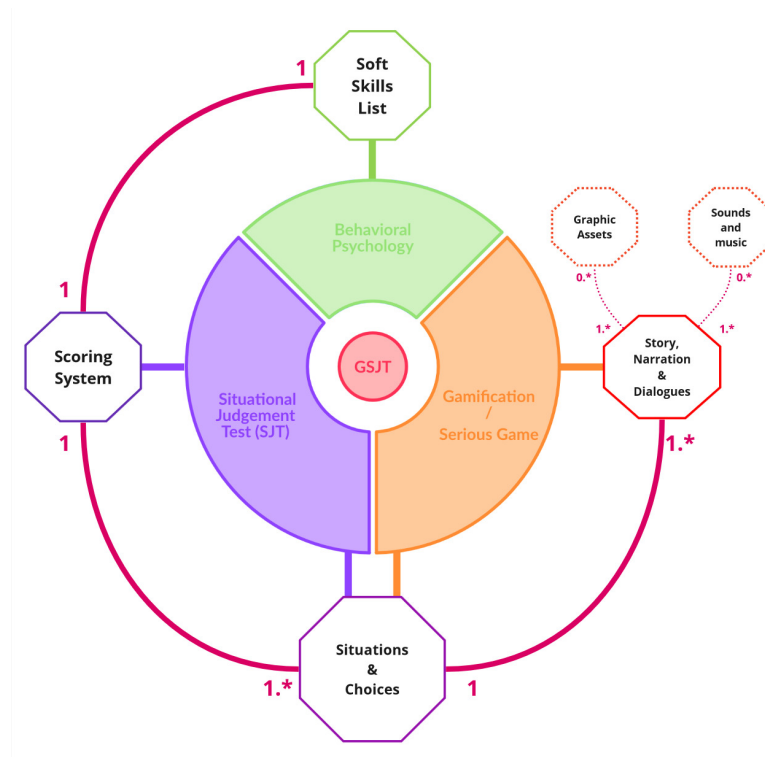


FIGURE 4.1: Gamified Situational Judgement Test Meta-Framework

Based on these dimensions, our proposed meta-framework for a Gamified Situational Judgement Test (GSJT) consists of **four core elements** designed to create a comprehensive evaluation tool for soft skills.

The first element is a rigorously defined **list of soft skills**, including cognitive abilities, personality traits, and transversal skills such as communication and teamwork. This list serves as the foundation for assessment, ensuring that the test is aligned with specific educational or occupational outcomes.

The second element is a well-designed **scoring system** aimed at quantitatively measuring an individual's soft skills based on their choices within various test scenarios. This system provides a nuanced yet straightforward way to analyze a test-taker's skill set.

The third element combines **situations and their alternative choices**, key facets of traditional Situational Judgement Tests (SJTs), into a unified component. Each situation and corresponding choice set is carefully designed to mirror real-life dilemmas, enhancing the ecological validity of the test.

The fourth and final element is the integration of **storytelling and dialogues** to turn these situations into interactive narratives, offering a more immersive and engaging test-taking experience with the aim of simulating contextual real scenarios.

In addition to these core elements, **secondary but non-mandatory elements** may also be introduced to enhance the immersion of the test experience. These could include graphic assets such as background imagery, character sprites, and audio assets like background music or sound effects. While these secondary elements are not essential for the core functionality of the GSJT, they could potentially highly elevate the user experience and thus impact the assessment's validity and reliability, as mentioned earlier.

However, it is important to note that this thesis does not aim to evaluate the impact of these secondary elements on the GSJT's reliability and validity. Exploring the implications of these secondary elements for assessment reliability and validity could be a fruitful avenue for future research.

Given the complexity and multi-domain nature of the GSJT meta-framework, a single-disciplinary approach is insufficient for capturing its nuanced requirements. Each dimension necessitates specialised expertise from psychology to game design and data analytics. Therefore, **a multi-disciplinary team comprising Psychologists, Human Resource Professionals, Assessment Experts, Data**

Analysts, Psychometricians, Game Designers, and Software Developers, among others, is essential for the effective conceptualisation, development, and implementation of a comprehensive Gamified Situational Judgement Test.

Hence, the meta-framework highlights the importance of a comprehensive and structured methodology that requires various actors' involvement in different stages of GSJT creation. The importance of a co-creative approach to GSJT development underscores the collaboration of different actors in this different area. Our meta-framework aims **to foster this collaboration**, which is crucial in generating **a more comprehensive, relevant, and engaging GSJT**. The findings presented in this section are informed by an amalgamation of literature research, field experiment insights, and analysis of practices from a video game company.

In conclusion, the proposed meta-framework offers a comprehensive approach for generating a GSJT, intending to streamline its design and production (see Figure 4.2).

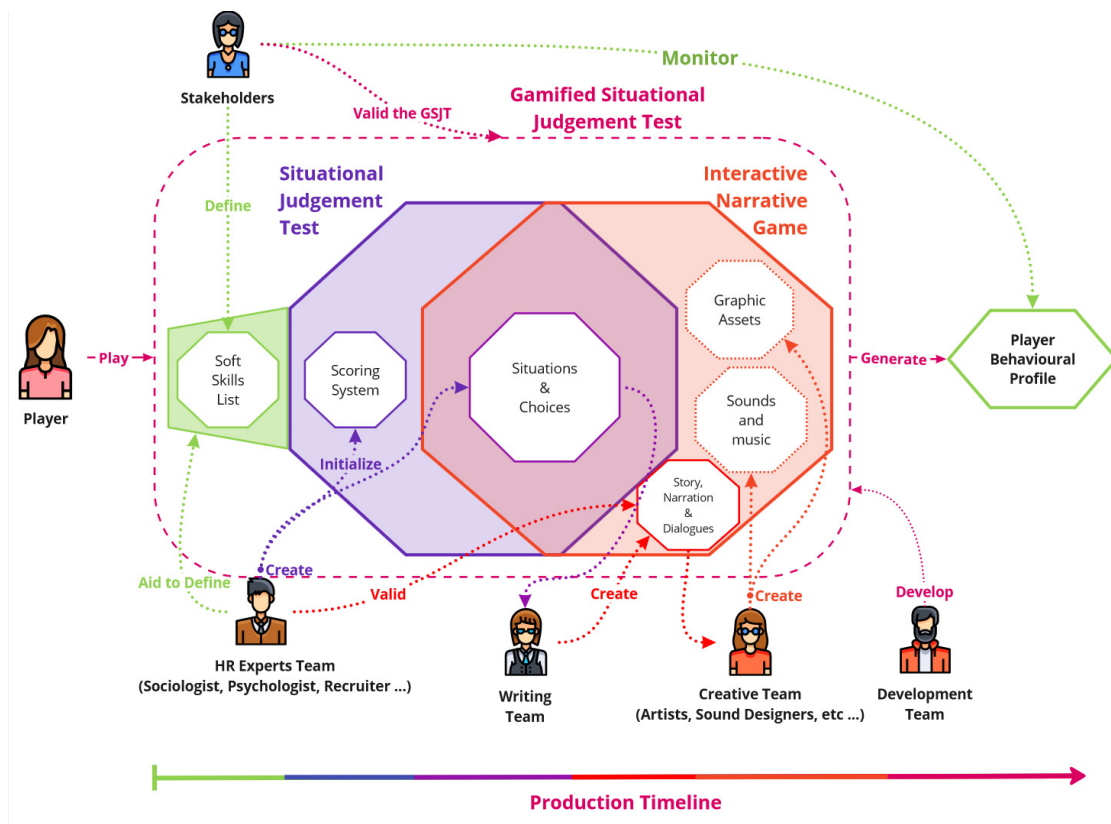


FIGURE 4.2: Gamify situational judgement test production timeline

Nevertheless, as highlighted by Bronsword's [215] study, creating video game development is complex and needs a multi-disciplinary team. Furthermore, within

the context of GSJT, the collaboration of domain experts with multi-disciplinary video game team development is essential. The expertise of these domain experts is instrumental in developing GSJT applications that encompass elements of challenge, entertainment, and success. Effective design strategies play a pivotal role in achieving these objectives, with the video game designer exerting a substantial influence on the overall development process. Hence, in the following section, we will propose our first framework based on the meta-framework to streamline the design and production of a GSJT by a multi-disciplinary team.

4.2 Toward a first framework: Identification and elaboration of the stages involved in the development of a GSJT

The overall approach that we propose revolves around the creation of a GSJT based on a branching narratives approach explained in section 3.3.3, which facilitates the construction of an entire GSJT through **five distinct processes** (see **Figure 4.3**): **(1) Defining the client request, (2) Generating a variety of traditional SJT situations, (3) Screenwriting for INGs, (4) Implementing the gamification based on ING, (5) Developing the final GSJT.**

As detailed in this section, these stages are organised in a coherent sequence. The difference between the meta-framework and the framework stages is relative to a deep exploration of each process. Finally, it is essential to note that a degree of flexibility is inherent to this approach, as it allows for iterative progress and revisions throughout the development process. Consequently, it is possible to navigate back and forth between phases or even to have certain phases unfold concurrently for different aspects of the GSJT. This adaptability ensures that our methodology can be effectively tailored to meet the unique requirements and objectives of various GSJT contexts.

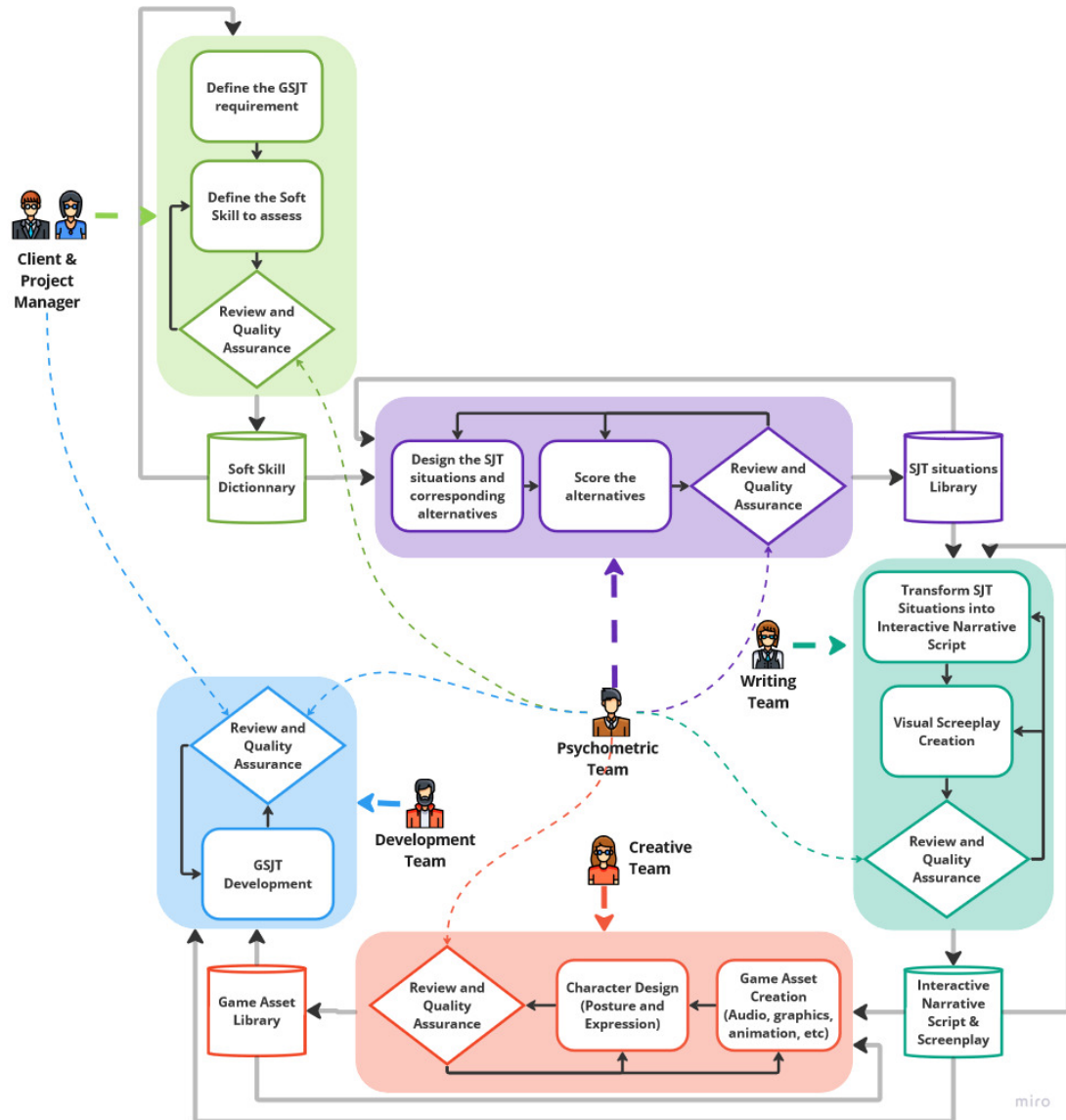


FIGURE 4.3: Gamified Situational Judgement Test Design Process in 5 Steps

4.2.1 Stage 1: Defining the client request

The initial phase in developing a GSJT to evaluate players' soft skills involves comprehensively delineating the client's specific request. This crucial process necessitates the involvement of **the project manager** and a thorough understanding of **the client's objectives and goals, achievable through collaborative dialogue, the collection of vital information, and an attentive examination of the client's needs.**

In this context, it is essential to consider any potential cultural, demographic, or

industry-related factors that may impact the test design and its efficacy. By gaining a lucid comprehension of the client's expectations and requirements, this stage lays the groundwork for the subsequent progression of the GSJT. Moreover, the knowledge obtained empowers the design team to align the anticipated outcomes with the test content (e.g., fantasy, realistic, or mixed setup), format (e.g., visual novels, point-and-click, interactive movie, or other), and difficulty level. This alignment ensures that the test is aptly tailored to the specific context.

Additionally, as the primary focus of assessment, it is imperative to accurately define the target soft skills for evaluation. Consequently, we recommend collaborating with the client and psychometric professionals to identify and describe the soft skills in succinct paragraphs. To facilitate the next phase of developing situational judgement test (SJT) scenarios for soft skills, we suggest providing concise examples of appropriate behaviour that one should demonstrate with the identified soft skills. The upcoming definition of the soft skill *Cooperation*, elaborated with a detailed protocol in section 4.3.1.1, illustrates a potential methodology for this process.

- **Soft skill Title:** *Cooperation*
- **Soft Skill Description:** *Enhancing interdisciplinary collaboration involves prioritising the overall well-being and client satisfaction to boost collective productivity. This requires moving beyond isolated strategies and promoting a collaborative mindset and conduct within your team and among your colleagues.*
- **Behaviour to Adopt:**
 - *Recognise and promote actions that support interdisciplinary cooperation, teamwork, and exchanging information and successful strategies.*
 - *Extend beyond your team's boundaries by empowering other teams or departments to utilise each employee's skills for diverse projects.*
 - *Establish collective goals alongside individual targets for a well-rounded approach.*

In conclusion, each identified soft skill will be catalogued in a soft skill library with the validation of the psychometrics experts for future project utilisation. Indeed, some clients may need help to articulate their specific soft skills. Thus, offering

an established soft skill dictionary can assist them in clarifying their requirements and ultimately improve the production workflow.

By accurately outlining the client's request, it not only establishes a streamlined and focused development process but also guarantees that the resulting GSJT coherently evaluates players' soft skills in the relevant context.

4.2.2 Stage 2: Generating various traditional SJT situations

The second stage in developing a GSJT for assessing players' soft skills involves the generation of various situational judgement test scenarios by the psychometrics team, composed of **HR experts, psychologists, and sociologists**. Leveraging their collective expertise and understanding of human behaviour, the team utilises the soft skills library and the specific soft skills requirements provided by the client **to create a diverse range of realistic and engaging situations that test the target soft skills.**

During this process, the psychometrics team carefully crafts each scenario to ensure it is relevant to the target population, industry, and cultural context. The team considers potential biases, aiming to create fair and inclusive situations that allow for accurate assessment of the soft skills in question. Additionally, each scenario is designed to incorporate varying complexity and challenge to accommodate test-takers with different proficiency levels in the assessed soft skills. For example, a situation that assesses *Cooperation soft skill* may be represented as follows:

SJT item that assesses Cooperation

You are part of a team working on a significant project that is vital to your organisation. One team member, although competent and enthusiastic, consistently ignores the established team processes and guidelines. This individual's actions have started to impact the progress of the project.

- A) Continue working and ignore their actions, focusing solely on your tasks. (1)
- B) Confront the individual directly and express your dissatisfaction about their disregard for the team's processes and guidelines. (2)
- C) Discuss the situation with the rest of the team without involving the individual in question. (2).
- D) Inform your team lead or manager about the situation. (3)
- E) Arrange a meeting with the individual to discuss the importance of adhering to team guidelines and explore ways to enhance their cooperation with the team.(5)

Therefore, the psychometric experts have **to generate a scoring grid** for each SJT item that will enable the generation of the player's behavioural profile. The psychometric experts may also provide more information about their scoring in the case of the GSJT being used as a formative assessment as the following table 4.1 depicts it.

To further enhance the coherence between the SJT and the final GSJT, we suggest that the psychometrics team collaborates closely with the client or the project manager throughout the review and quality assurance process. This collaborative approach fosters an open exchange of ideas, feedback, and insights, ensuring that the resulting SJT scenarios are well-aligned with the client's objectives and expectations.

This phase forms the basis for a **comprehensive and practical GSJT by generating diverse SJT scenarios** that accurately evaluate the target soft skills within the given context. Under the supervision of the psychometrics team, the validated situations can be catalogued in an SJT situations library for future use.

Choice	Score	Remarks
A	1	This option is the least cooperative as it disregards the issue and can potentially harm the project.
B	2	Confrontation might not encourage cooperation and may lead to conflict within the team.
C	2	While this choice brings the issue to attention, it might create a negative environment and be perceived as gossip.
D	3	Involving a team lead or manager can be a feasible choice if other resolution attempts have failed or the behaviour seriously impacts the project.
E	5	This is the most cooperative approach, as it directly addresses the issue, underlines the importance of teamwork, and seeks a constructive solution beneficial for the whole team.

TABLE 4.1: Cooperation Item's Answer Scoring Grid

This thorough and cooperative method supports the creation of a robust and significant assessment tool for appraising players' soft skills and streamlining the production workflow.

4.2.3 Stage 3: Screenwriting for INGs

The third stage in developing a GSJT for evaluating players' soft skills focuses on **scriptwriting and screenwriting for an interactive narrative game**. This stage engages **a writing team of game designers, level designers, and screenwriters** with expertise in crafting immersive and engaging game experiences. Drawing from the SJT library of situations created in the previous stage, the storytelling component presents scenarios that challenge players to apply their soft skills in-game situations.

Initially, the writing team produces a script's first draft, which features characters' dialogues and allows players to assume their roles in the game. This script is designed within the framework of the setup defined by the client during *Stage 1*. Subsequently, the team creates a storyboard that visually represents the narrative flow and provides a valuable reference for the upcoming stages of development.

As the script and storyboard are developed, the writing team pays close attention to the narrative structure and pacing, ensuring the game experience remains engaging and immersive. Moreover, the team considers various player choices and

potential branching narratives, creating a dynamic environment that adapts to players' decisions and provides a more accurate assessment of their soft skills.

Finally, the psychometric team will validate the various scripts and screenplays to collect them into an interactive narrative script and screenplay library for reuse purposes.

In summary, *Stage 3* combines the expertise of game designers, level designers, and screenwriters to develop scripts and storyboards for an interactive narrative game that effectively assesses players' soft skills within the specified context. By leveraging the SJT library of situations and adhering to the client's setup, this stage contributes to **the gamification design of the SJT**.

4.2.4 Stage 4: Crafting the gamification based on ING

The fourth stage in developing a GSJT for evaluating players' soft skills involves **implementing game elements**. A creative team consisting of **game artists, audio managers, animators, VFX artists, and other specialists** is assembled to produce various game assets and character designs based on the script and screenplay provided in the previous stage.

The primary objective of this stage is **to introduce engaging and immersive aspects of an interactive narrative game** within the context of soft skill assessment. The creative team collaborates closely with the writing team to ensure that the game assets, character designs, and audiovisual elements align with and enhance the narrative structure and player experience. This collaborative approach helps create a cohesive game environment that effectively combines the immersive aspects of interactive narrative games to assess players' soft skills.

During this stage, the creative team crafts game assets that enrich the interactive narrative experience. These assets may include detailed environments, props, music and audio, and graphics that support the story and provide context for the soft skill challenges. Additionally, they create character designs and animations that bring the narrative to life, ensuring players can easily relate to and empathise with the characters they encounter. Sound design and music are also carefully curated to complement the game's atmosphere and immerse players in the narrative.

In section 3.3.3.2, we highlighted the importance of realistic character design and interaction in an interactive narrative game for soft skill assessment. Since soft skills are often transversal and interpersonal, the game should feature characters that **closely resemble real-life individuals in appearance and behaviour**, allowing players to immerse themselves and elicit accurate responses when facing soft skill challenges. Believable characters with nuanced expressions, body language, and dialogue help simulate real-world interactions and promote developing and demonstrating the target's soft skills.

Additionally, realistic character interactions enhance the credibility and reliability of the assessment tool, as the GSJT aims to simulate realistic interpersonal situations. Thus, **realistic character design and interaction is an essential aspect of gamification for an effective and valid soft skill assessment tool**.

In the end, the psychometric group will closely observe the various character designs to confirm that the artists accurately represent the emotions conveyed by the scriptwriters. Subsequently, the assorted game assets will be methodically collected and compiled into a comprehensive game asset library, guaranteeing their reuse in upcoming projects.

Stage 4 encompasses the expertise of a **creative team to implement game elements** within an interactive narrative game designed for soft skill assessment. By developing engaging game assets, character designs, and audiovisual components based on the script and screenplay from the previous stage, this phase contributes **to creating an immersive and captivating assessment tool with the aim to provide an enjoyable gaming experience**.

4.2.5 Stage 5: Implementing the GSJT

The primary objective of this stage is **to create a seamless and engaging gaming experience that integrates the immersive aspects of interactive narrative games with the assessment of players' soft skills**.

Throughout this stage, the development team closely collaborates with the psychometric team to ensure that the narrative situations remain consistent with the

overall vision and objectives of the GSJT. This collaboration is essential for maintaining the integrity of the soft skill assessment tool and ensuring that the final product effectively evaluates players' soft skills within the specified context.

The development team utilises the narrative script and screenplay to guide the construction of the game environment and the implementation of the interactive elements. They also incorporate the game assets, such as character designs, animations, and audiovisual components, to create a cohesive and visually appealing gaming experience as depicted in the screenplay.

Finally, under the supervision of the psychometric team, this last stage is instrumental in building the GSJT by combining the narrative situations. This process creates a tailored, engaging, and immersive interactive narrative game designed for soft skill assessment. By integrating the work from previous stages, **the final product could effectively evaluate players' soft skills while providing an enjoyable and contextually relevant gaming experience.**

4.2.6 Conclusion

In this section, we described, on the one hand, **a three-step meta-framework to create a GSJT**. This meta-framework aims to guide GSJT designers in conceiving their framework or GSJT. Subsequently, we proposed **our framework to generate a GSJT based on INGs**. The framework emphasises collaboration and teamwork while highlighting the importance of psychometrics experts through the whole process of design and conception. In the following section, we will experiment with our framework by implementing a GSJT.

4.3 Experimentation 1: A first GSJT

The experimentation of our framework is divided into two stages. The first stage of the experimentation is **implementing our framework into a GSJT called *Flight***. The second stage aims **to validate the proposed GSJT** by conducting an online experiment with a diverse group of participants who will play the GSJT.

4.3.1 Stage 1: Experiment Objectives

The primary objective of the first stage is **to provide preliminary evidence supporting the validity of the proposed initial framework in a real-world context** by scrutinising the design of a GSJT. Consequently, an hypothesis concerning the design of a GSJT were tackled in this stage before presenting the resulting GSJT to the participants for *phase 2*. The hypothesis to be confirmed is the following:

- **Can the development and application of a meta-framework and a framework facilitate the establishment of specialized frameworks that provide a clear delineation of roles and responsibilities of all parties involved in the design and production process? This approach may enhance clarity and effectiveness in the game-based behavioural recognition process.**

As explained in Chapter 3, the creation of GSJT is recent, and to our knowledge, no framework has been developed for this specific context. Hence, this study was essential in collecting insight into GSJT Designers' working patterns and personal design strategies.

4.3.1.1 Stage 1: Experimental Methodology

We created a GSJT using the proposed framework to compare the time taken and the number of people involved in generating the GSJT. Due to the client's confidentiality reasons concerning the creation processes, an extensive detailed creation process cannot be provided in this thesis. However, after conducting a post-mortem examination of the project through several briefings with the GSJT designers, we identified the challenges the GSJT designers encountered throughout the development process.

4.3.1.2 Stage1: Experimental GSJT design and development

The creation and development of the experimental GSJT design involved the collaboration of one entity and three individuals:

- **The client** - An international banking group sought a game-based assessment to evaluate their managers' leadership and management soft skills. They also desired direct and delayed feedback integration for learning purposes.
- **A product owner** - This individual also served as a psychometric expert and writer during the development of the GSJT.
- **A gamification expert** - Responsible for tasks such as screenwriting, sound design, graphic design, and development within the GSJT.
- **A game artist** - Focused primarily on generating 2D game assets, including background and character designs for the GSJT.

The client's goal was to develop a game with an average **duration of fifteen minutes**, capable of assessing multiple soft skills, but also initiate the players to the concepts of the soft skills. After establishing an initial framework, the client and product owner identified **nine soft skills**. These sessions helped clarify the soft skills the client wanted to assess using the game, as understood by the psychometric expert. An **interview with the expert** revealed that they adhered to the framework by conceptualising various scenarios with alternatives where soft skills could be evaluated. **The expert collaborated with the clients in workshops** to ensure the scenarios and alternatives aligned with the client's objectives. Subsequently, **the client and psychometric experts assigned two types of scoring to each item alternative**. First, the client wanted to **track the player's learning progress** by employing a **player score** that the client instantiated. Second, the psychometric expert evaluated each item alternative's soft skill. This **secondary scoring system will monitor the player's behavioural profile** and any changes that may occur throughout the game (see Figure 4.4).

In the second stage, the script was formatted similarly to a **theater script**, which allowed the screenwriters to create a **visual storyboard** based on the text. This storyboard enabled the gamification expert to develop the GSJT and communicate with the game artist regarding the required game assets.

Ultimately, **the decision tree representing the game consists of seven narrative modules**, each featuring **three or four alternatives**. Each node of the tree represents a **choice made** in a situation inside the initial situational

judgement test, and each node will fluctuate **the behavioural profile** using a corresponding scoring matrix.

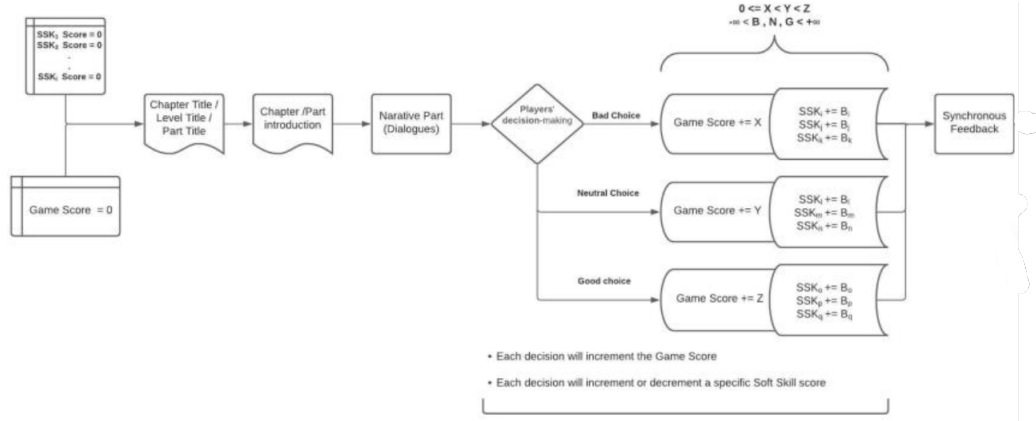


FIGURE 4.4: Narrative Module Internal Structure Example

In addition, to emulate real-life situations where critical decisions can have significant consequences, the game's fifth scenario may be the player's final choice, leading to the game's conclusion if the player selects one of the incorrect alternatives. The following section presents an online experiment conducted with the created GSJT, followed by a survey (see Appendix 7) that also serves as a debrief for players.

4.3.1.3 Stage 1: Results

The creation of the GSJT took approximately **two months**, with one month focused on creating the interactive and narrative situations and script and one month on creating the production (development and game assets creation). The client was satisfied with the first whole result and the creation (see Figure 4.5) and proposed only some graphics and minor textual modifications.

On the other hand, we debriefed in a post-mortem interview with the GSJT designers about the conception of the GSJT. Firstly, the GSJT designers and development teams were satisfied with the framework for several reasons:

- **Structured item approach:** The framework offered a well-defined and coherent structure for the entire design and development process. This organisation enabled the team to plan, design, and implement the gamified situational judgement test methodically.

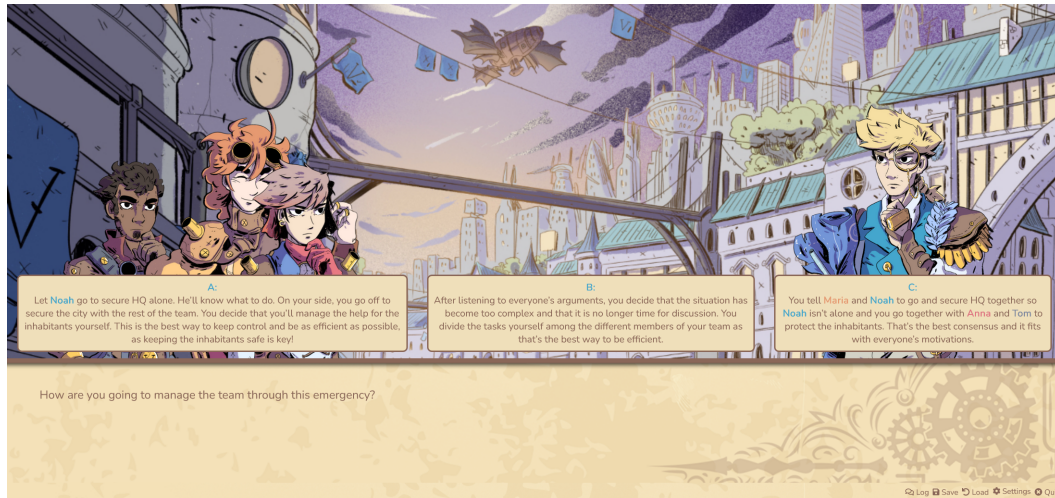


FIGURE 4.5: Screenshot of the Result GSJT

- **Collaboration enhancement:** Despite the lack of comparison methodology, the GSJT designers emit that the framework facilitated better communication and collaboration among team members, including subject matter experts, designers, and developers. It encouraged open discussions and exchange of ideas, potentially leading to a more effective and engaging GSJT.
- **Cost efficiency:** By following the structured framework, the team could optimise the use of resources and reduce the likelihood of costly errors or rework, making the development process efficient.

However, one aspect of the framework that proved difficult for the GSJT designers and teams was the time-consuming nature of the research and content development phase. Gathering information on the specific job context and developing realistic scenarios required significant effort and collaboration with subject matter experts. This challenge was addressed by allocating additional resources and time to this phase with a workshop with the psychometric expert and the human resources people of the client and by emphasising the importance of thorough research and content development to ensure the quality of the GSJT. In addition, the psychometric expert was alone to produce the scoring system and emit that the need for external help would enhance the construct validity of the GSJT.

Overall, the satisfaction expressed by the GSJT designers and development teams with the framework shows **preliminary results in its potential for enhancing collaboration and cost-efficiency in creating gamified situational judgement tests.**

4.3.2 Stage 2: Experiment Objectives

The primary objective of the second stage is to provide **priliminary evidence from the player that such gamified situational judgement tests are coherent in their purposes to analyse soft skills**. Consequently, several hypotheses concerning the coherence of the GSJT were tackled in this stage. The hypotheses to be confirmed include the following:

- **Is our game-based approach model proposal enabling the generation of an immersive serious game to recognise soft skills for recruitment and selection process.**
- **The application of computational psychometrics may assist psychometric experts in developing a relevant in-game soft skill measurement system.**

As explained in Chapter 1, evaluating soft skills is a complex and challenging endeavour due to their intangible, subjective nature. Unlike hard skills, which can be quantified through standardised tests and certifications, soft skills lack universally accepted metrics, making their assessment more difficult [9].

Hence, this experiment stage played a vital role in collecting preliminary insight on the coherence for the player to be evaluated on their soft skills through a GSJT. Through analysing this data, we aim to find evidence that GSJT may be used for soft skills assessment.

4.3.2.1 Stage 2: Participants

Among sixty-five players, **forty-eight** (twenty-three females, twenty-three males and two undefined aged between **18 and 75**) answered the self-assessment survey. The majority of the participants (**66%**) came from professional backgrounds, with **fourteen holding managerial positions**. A diverse set of **eleven industries** were represented within the participant pool. Interestingly, **seven of the players identified themselves as non-gamers**, providing valuable perspectives from individuals with varying levels of gaming experience.

4.3.2.2 Procedure

This experiment was conducted online, involving five participants who partook in a live session with an instructor. At the same time, the remaining sixty players engaged individually with only the in-game tutorial for guidance and a self-assessment survey for debriefing. The live session began with the instructor offering brief gameplay instructions and clarifying the game's learning and assessment objectives. Subsequently, **a post-game debriefing was held with the group of five.**

The survey's primary purpose was to determine whether the game was **engaging and user-friendly for the players**. Additionally, the survey aimed to ascertain if the players believed the GSJT was **a suitable and coherent method for assessing soft skills in a game-based environment**.

To achieve these goals, the survey included questions about the player's overall experience, the ease of use and understanding of the GSJT, and their perception of the game's effectiveness in assessing soft skills (see Appendix 7).

4.3.2.3 Stage 2: Results and discussion

The survey results demonstrated that **all participants in the experiment found the gameplay straightforward and user-friendly**, eliminating the need for an instructor's guidance. The GSJT received favourable feedback, with players' average game satisfaction score being **7.9 on a scale of one to ten**. Participants universally commended **the gamification elements, including scenarios, graphic assets, and sound design, for their immersive and engaging nature**. This immersive experience was further amplified by incorporating realistic situations into a fantastical setting, as reported by live session players and **90.6% of survey respondents who had encountered similar situations in their professional or academic lives**.

Nonetheless, **the user interface could use some enhancements**, as the live session revealed particular challenges encountered by players. These include revisiting previous scenes in the interactive narrative game or reviewing the choices they have made throughout the game. Both the live session and survey results underscored the importance of debriefing with an instructor to validate the learning objectives of the GSJT fully.

The survey also investigated whether the GSJT **could impact players' future behaviours** and enhance their work performance as a training tool. A positive response was given as **71.9% of players believed in its ability to alter their future behaviours, while 90.6% thought it could enhance their work performance**. The experiment disclosed that eight out of fifteen players who claimed to know at least **80% of the evaluated soft skill meanings before the game achieved the highest or a high game score**. In comparison, the remaining players obtained a medium score. However, incorporating learning and assessment objectives within the same game could bias the results in terms of behavioural assessment. Consequently, this research focuses on gathering preliminary insights that a GSJT developed with our framework **could potentially recognise behavioural** patterns and profiles rather than examining the game's validity and reliability in assessing soft skills.

Therefore, we perform clustering analysis with a subset of players (48) who completed the game with the maximum number of possible choices (7). We performed a **cluster analysis for each situation**, allowing for the study of clustering evolution throughout the game. This analysis identified **three clusters** and revealed that an individual's membership would likely stay the same after the fourth choice. Only **three of the forty-eight players (27, 32, 38 see Figure 4.6)** experienced a change in their cluster affiliation after the fourth situation. This finding indicates a high predictive value from the first four-game situations, enabling a second cluster analysis of the dataset, including players who reached a premature end of the game (incorrect choice in the fifth situation).

The second cluster analysis (see Figure 4.7) was conducted on all players (**65 players**). We observed a pattern with the individuals of the cluster 1 tending to have higher results in every soft skill, followed by the individuals of the cluster 2 and 0. This pattern may be attributed to selecting several leadership soft skills (see Figure 4.8), which are semantically related.

However, figure 4.8 unveils the existence of **macro-SSKs**, such as leadership and trustworthiness, in our experiment. We define a macro-SSK as **a soft skill that encompasses or is strongly associated with other soft skills**. Our scoring model revealed that the "*Leadership Soft Skill*" and the "*Trustworthiness Soft Skill*" have higher maximum scores due to their ubiquitous presence in item responses and the ratio of links between all SSks. Further studies are required to understand the nature and impact of the connections between soft skills. The

Player	Choice 1	Choice 2	choice 3	Choice 4	Choice 5	Choice 6	Choice 7
#1	0	1	2	2	2	2	2
#2	0	2	2	1	1	1	1
#3	0	1	1	1	1	1	1
#4	2	2	2	1	1	1	1
#5	1	1	0	0	0	0	0
#6	1	1	0	0	0	0	0
#7	1	1	0	0	0	0	0
#8	0	0	0	2	2	2	2
#9	0	0	2	2	2	2	2
#10	0	0	0	0	0	0	0
#11	0	1	2	1	1	1	1
#12	0	1	2	1	1	1	1
#13	0	1	2	2	2	2	2
#14	0	0	0	0	0	0	0
#15	0	0	0	0	0	0	0
#16	2	2	2	2	2	2	2
#17	2	2	2	1	1	1	1
#18	0	1	1	1	1	1	1
#19	0	0	0	0	0	0	0
#20	0	0	0	0	0	0	0
#21	0	0	0	0	0	0	0
#22	0	1	1	1	1	1	1
#23	0	0	0	0	0	0	0
#24	1	1	1	2	2	2	2
#25	0	0	2	1	1	1	1
#26	0	1	2	2	2	2	2
#27	0	0	0	0	0	0	0
#28	0	2	0	1	1	1	2
#29	1	1	0	0	0	0	0
#30	0	0	0	0	0	0	0
#31	0	0	0	0	0	0	0
#32	0	0	0	0	0	0	0
#33	1	1	0	1	1	1	2
#34	0	0	0	0	0	0	0
#35	0	0	0	2	2	2	2
#36	0	0	0	0	0	0	0
#37	0	0	0	0	0	0	0
#38	0	0	0	0	0	0	0
#39	0	0	0	1	1	1	2
#40	0	0	0	0	0	0	0
#41	0	1	2	2	2	2	2
#42	0	0	0	0	0	0	0
#43	1	1	2	1	1	1	1
#44	0	0	0	0	0	0	0
#45	0	0	0	0	0	0	0
#46	1	1	0	0	0	0	0
#47	0	0	0	0	0	0	0
#48	0	0	0	0	0	0	0

FIGURE 4.6: Clustering evolution of the forty-eight players

assessment objectives and the assessment model's structure also account for the results in figure 4.8. The assessment objectives direct the GSJT design to assess soft skills closely related to leadership and management. This focus defines **a referential behavioural profile (RBP) for stakeholders**, consisting of the nine soft skills generated through the framework. In the context of company recruitment or training, this RBP will be compared to the player's behavioural profile generated after the game. In other words, members of cluster 1 are more likely to align closely with or even match the RBP, as demonstrated in our experiment.

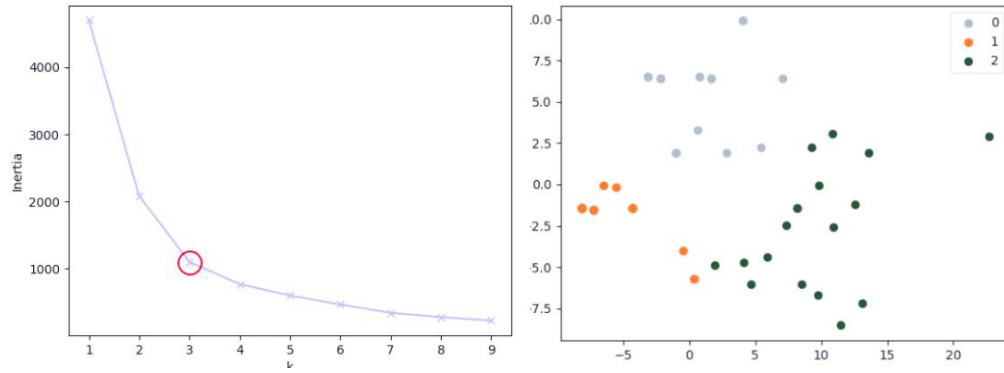


FIGURE 4.7: Elbow method and K-means clustering on the sixty-five players after the fifth game situation

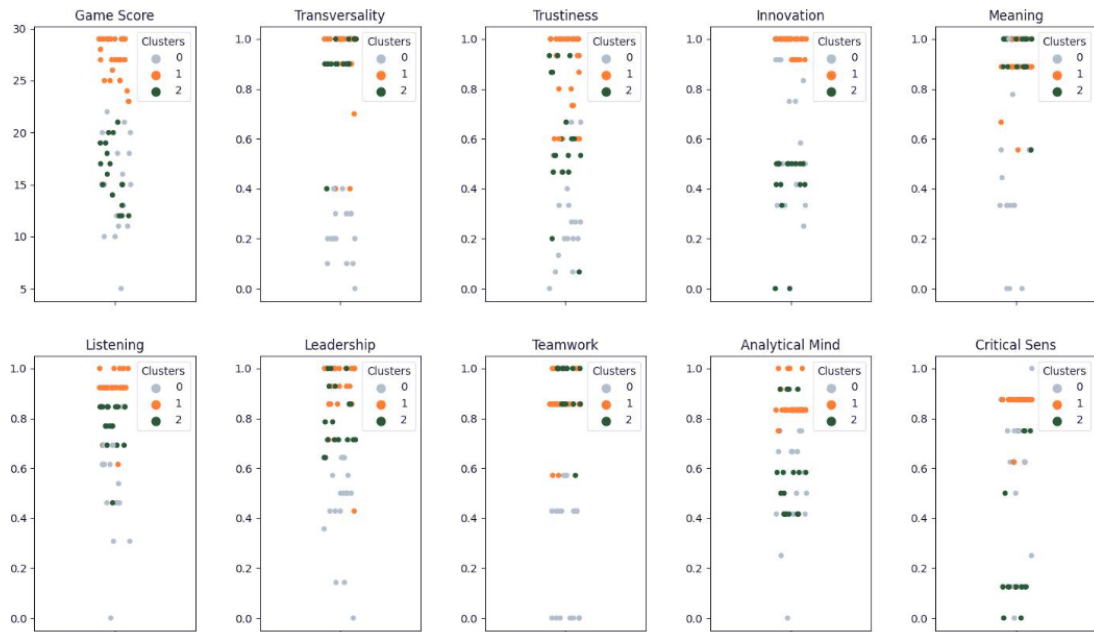


FIGURE 4.8: Players distribution by game score and soft skills

To summarise, the GSJT was well-received by participants, who found it immersive and engaging. While the user interface could be improved and debriefing with an instructor is essential, the GSJT shows promise in its ability to impact players' future behaviours and enhance their work performance.

The clustering analysis provided valuable insights into players' behavioural patterns and profiles. By examining the data, it is evident that **macro-SSKs, such as leadership and trustworthiness**, play a significant role in players' performance. Moreover, the analysis revealed the importance of **referential behavioural profiles in the context of company recruitment and training**.

In conclusion, this research presents **preliminary results that a GSJT developed using our framework has the potential to recognise and evaluate behavioural patterns and profiles.** The framework and the experimentation have been the subject of a peer-reviewed scientific paper [\[216\]](#). Yet, further studies are needed to explore the nature and impact of the relationships between soft skills and better understand these tools' applicability in various professional and academic settings.

Chapter Overview

This chapter provides insight into **GSJT creation**, emphasizing the development of a **Meta-Framework** and a **Preliminary Framework**, designed to **aid GSJT designers** in the entire process. We introduced the Meta-Framework in **three steps**:

- **Step 1:** Defining the soft skills and the client needs
- **Step 2:** Designing the Situational Judgement Test
- **Step 3:** Gamifying the Situational Judgement Test

The Meta-Framework provides a high-level abstract structure for GSJT creation. Meanwhile, the Preliminary Framework not only details stakeholder roles in GSJT development but also provides a more detailed exploration of the GSJT creation process in **five stages**.

- **Stage 1:** Characterization of the client's requirements
- **Stage 2:** Generation of various SJT scenarios
- **Stage 3:** Crafting the narratives for INGs
- **Stage 4:** Implementation of gamification elements
- **Stage 5:** Development and finalization of the GSJT

Subsequently, we navigate through two experimental phases, each purposed with distinct objectives. The initial experiment focuses on **validating the design and development process of the GSJT frameworks**. The second implements the GSJT with chosen participants, details procedures, analyses results, and includes a thorough discussion about **the use of such assessment means for soft skill recognition**.

This fourth chapter lays the groundwork for accomplishing our **research objectives A and B**. We probe into the potential assessment capabilities of the GSJT models and provide preliminary assistance in constructing such evaluative tools.

Chapter 5

Proposing a Second Framework based on a Modular Model, Conjugated with a Drama Manager

In the previous chapter, we constructed a **framework facilitating the development of GSJT**, premised on a branching narrative system. This structure, enabling the player choices to fork and weave across an intricate narrative network, instigates a wealth of complexities despite its simplicity of comprehension for GSJT designers.

The branching narrative system is laden with seven significant challenges. As the number of narrative branches proliferates, **the intricacies of supervising and ensuring narrative coherence across each strand increase exponentially**, thereby escalating the resource demands considerably.

Secondly, fabricating a unified narrative encompassing all potential player paths becomes complicated due to the mandate to assess each player on all the soft skills stipulated by the clients. A case in point is the creation of a GSJT that aspires to be fair in evaluating soft skills. It calls for a condition where every player is exposed to **an equivalent array of scenarios that test a particular soft skill**. However, within the constraints of a branching system, this endeavour might be highly demanding as the narrative intricacies of the GSJT may need to be more readily aligned with its evaluative intent.

The third challenge lies in striking **a delicate balance between player agency and authorial control**. The conjunction of a branching system and the evaluative goals of a GSJT could compel designers to guide players along a predetermined path, thereby stifling the potential of player agency within the GSJT. This could undermine the dynamic autonomy of the player, lessening the game's interactivity.

Fourthly, **the likelihood of every player exploring all narrative branches could be higher**, resulting in a considerable proportion of underutilized content. Fifthly, **maintaining consistent character development across varying paths presents a strenuous task**. In sixth place, the augmented replayability the branching narrative provides carries an inherent **risk of player fatigue**. The seventh and final challenge involves ensuring the absence of bugs and glitches across every narrative pathway, necessitating complex, **extensive quality assurance testing**.

Given the complexities and nuances of the branching narrative system, this chapter introduces **an innovative framework devised to surmount these challenges**. The aim is to heighten both the interactive gaming experience and the precision of the soft skills assessment in GSJT.

As a response, this chapter will introduce a GSJT based on a **semi-emergent modular model**. This model aims to manage branching narratives efficiently, thus reducing the complexity of maintaining various narrative paths, enhancing narrative coherence, and striking a judicious balance between player agency and authorial control with the GSJT soft skill recognition objective.

Additionally, this chapter will also delve into **the exploration of narrative coherence and the dynamic tension between player agency and authorial control**. To mitigate these challenges, we propose a contextually adapted **drama manager**. This solution allows a delicate equilibrium of player agency and authorial control. It ensures narrative coherence, thereby enhancing the player's immersive experience and satisfaction without impeding the soft skill recognition objective of the GSJT.

5.1 Semi-Emergent Modular Gamified Situational Judgement Test Approach

In contrast to existing GSJT literature and our first framework, which suggests the development of multiple branches following each player choice in order to generate a comprehensive player storyline and profile throughout the game, our novel approach involves dividing each SJT situation and its associated gamified story into separate **narrative modules** (refer to Figure 5.1). In the subsequent subsection, we will delve into the concept of narrative modules and provide a detailed explanation of their significance and function.

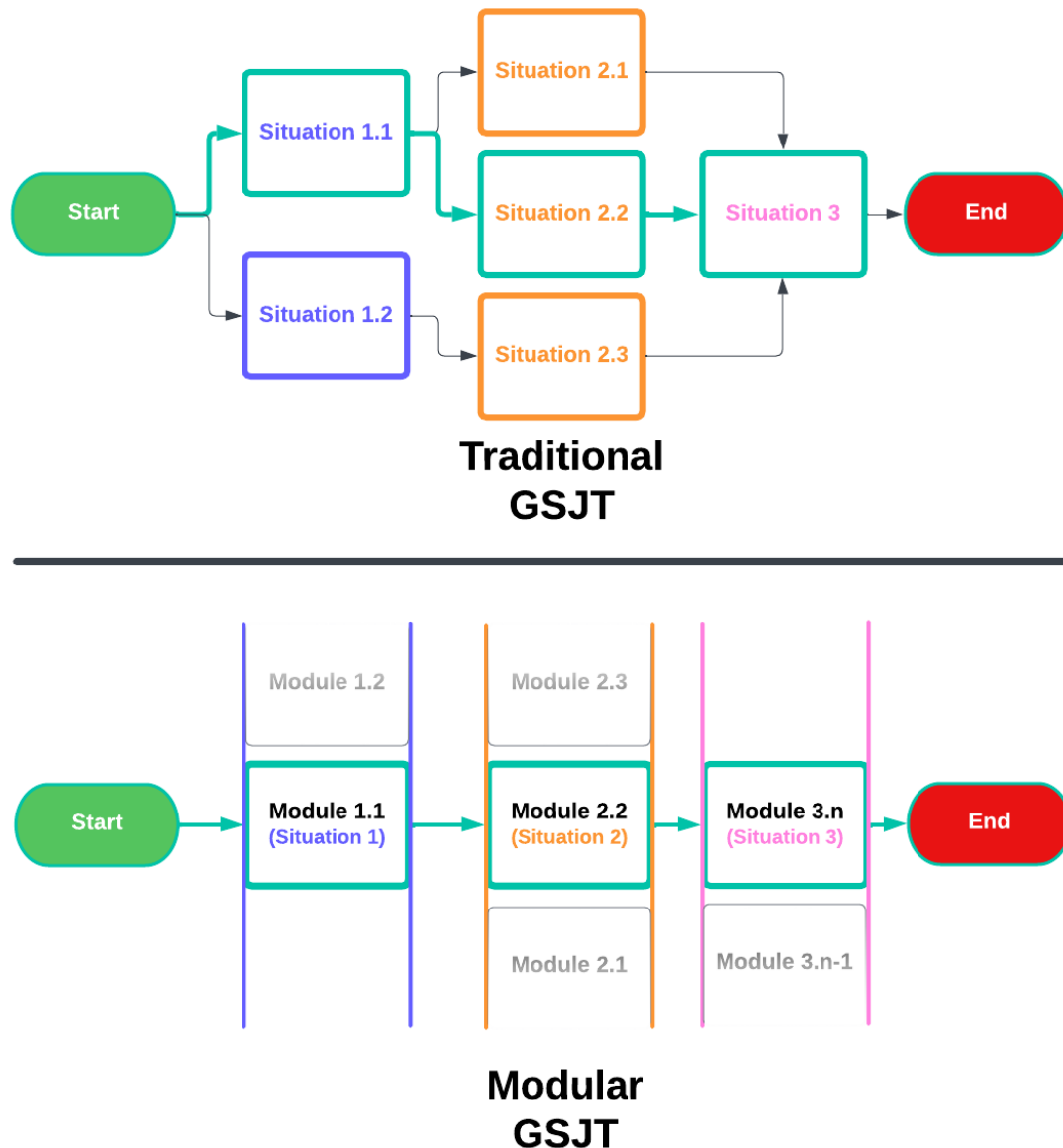


FIGURE 5.1: Traditional and Modular Gamify Situational Judgement Test

Drawing inspiration from Cameron’s work [217], narrative modules serve as the foundational elements that create a more comprehensive narrative framework. In this second proposed model, these distinct, **self-sufficient narrative segments can operate independently within the larger context**. This is achieved by integrating multi-branching tree storytelling structures with a modular approach.

5.1.1 Narrative modules

We consider a **narrative module as a distinct portion of the GSJT** that encapsulates a specific narrative event or sequence of events. These modules are designed to provide players with an engaging story experience and to allow them to influence the course of the narrative through their actions and decisions.

The narrative module is an individual building block in the larger narrative structure of the game. Each module may contain characters, settings, plot elements, and possible outcomes. However, it is also connected to other modules in the game’s overall narrative network through **interfaces** (see Subsection 5.1.2). A player’s progression from one module to another can be influenced by various factors, such as their in-game decisions and success in completing tasks or challenges.

In more technical terms, a narrative module in a GSJT could be a combination of scripted events, dialogues, and environmental storytelling elements that are organised in such a way that they can adapt and respond to the player’s actions. This is often facilitated by underlying algorithms and game mechanics, such as players’ inputs and choices that monitor the player’s actions and adjust the game’s narrative accordingly.

It should be noted that the concept and implementation of narrative modules can vary significantly depending on the specific GSJT and the creative vision of its designers. Some GSJT might utilize a highly modular narrative structure with many possible paths and outcomes. In contrast, others might employ a more linear narrative with fewer branches but more depth in each module. In either case, the goal of the narrative module is to provide a compelling and interactive narrative experience for the player that can be replayed, as the story will differ for each player’s session.

5.1.2 Interfaces

Within the realm of modularity, **interfaces** hold significant importance in delineating how modular components can be assembled to establish more expansive systems (see Figure 5.6). In the context of our model, the GSJT designers can merge modules using interfaces that facilitate seamless communication and connection without disrupting the narrative flow.

Therefore, in a GSJT structure where the player progresses through a series of narrative modules based on their decisions, the interfaces between modules would determine how player choices affect the transition from one narrative module to another.

The interfaces between these narrative modules might be defined as follows:

- **Module Transition Interface:** This would define how the game transitions from one narrative module to another. This interface could include methods for determining which module to transition to based on the player's decision history and soft skills assessment history, as well as methods for initializing the new module and terminating the old one, such as predetermining dialogues or cut-scenes. For instance, if a player completes a narrative module set within a forest, the subsequent module should also commence within a forest to maintain narrative coherence (see Figure 5.2).
- **Decision Impact Interface:** This would define how a player's decisions within a module impact future modules. This could include methods for tracking the effects of decisions and making this data available to future modules. For instance, if a player decides to fire a member, whom we will name '*Thrall*', of his team or party, the next module will consider this state and remove the path in the following module where '*Thrall*' may still be in the party. Future dialogues in the following modules may also reflect the previous decision to make the player feel that their decisions have meaningful impacts on the unfolding narrative.

Then, our model proposes using predetermined modules where the story is created through the game without a predetermined story or narratives. The choice of the following narrative modules is emergent, hence proposing a **unique story for the player**.

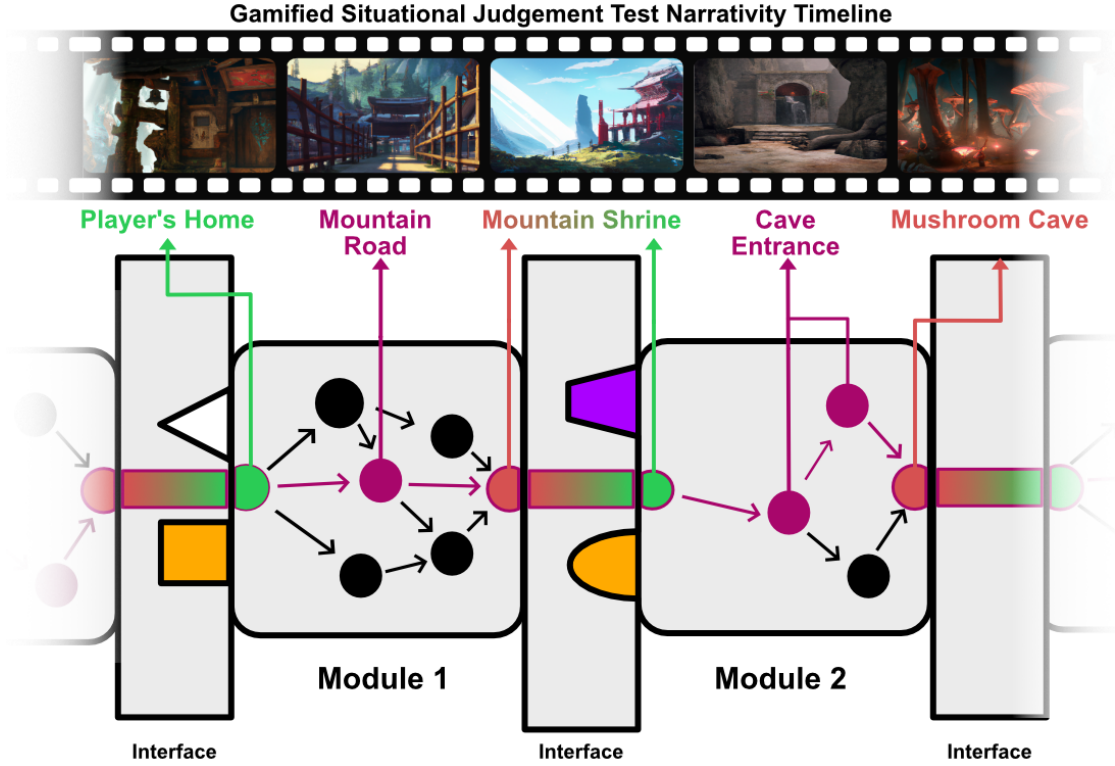


FIGURE 5.2: Narrative Modular combination

In summary, interfaces define the expected interactions between different modules in the context of modularity. They allow the modules to operate independently while enabling necessary communication between them. The design of these interfaces would depend on the specifics of the game's narrative structure and the level of complexity in the player's decision-making process. However, our goal would be to create a seamless, engaging experience where the player feels their decisions have meaningful impacts on the unfolding narrative.

5.2 Global Semi-Emergent Modular GSJT Design Approach

5.2.1 GSJT stakeholders' and Roles Identification

Expanding on our initial framework (see section 4), this novel framework is focused on **the iterative generation of several narrative modules**. Therefore, this second framework is focused on creating each narrative module by following the same steps of the preliminary framework (see Figure 5.3).

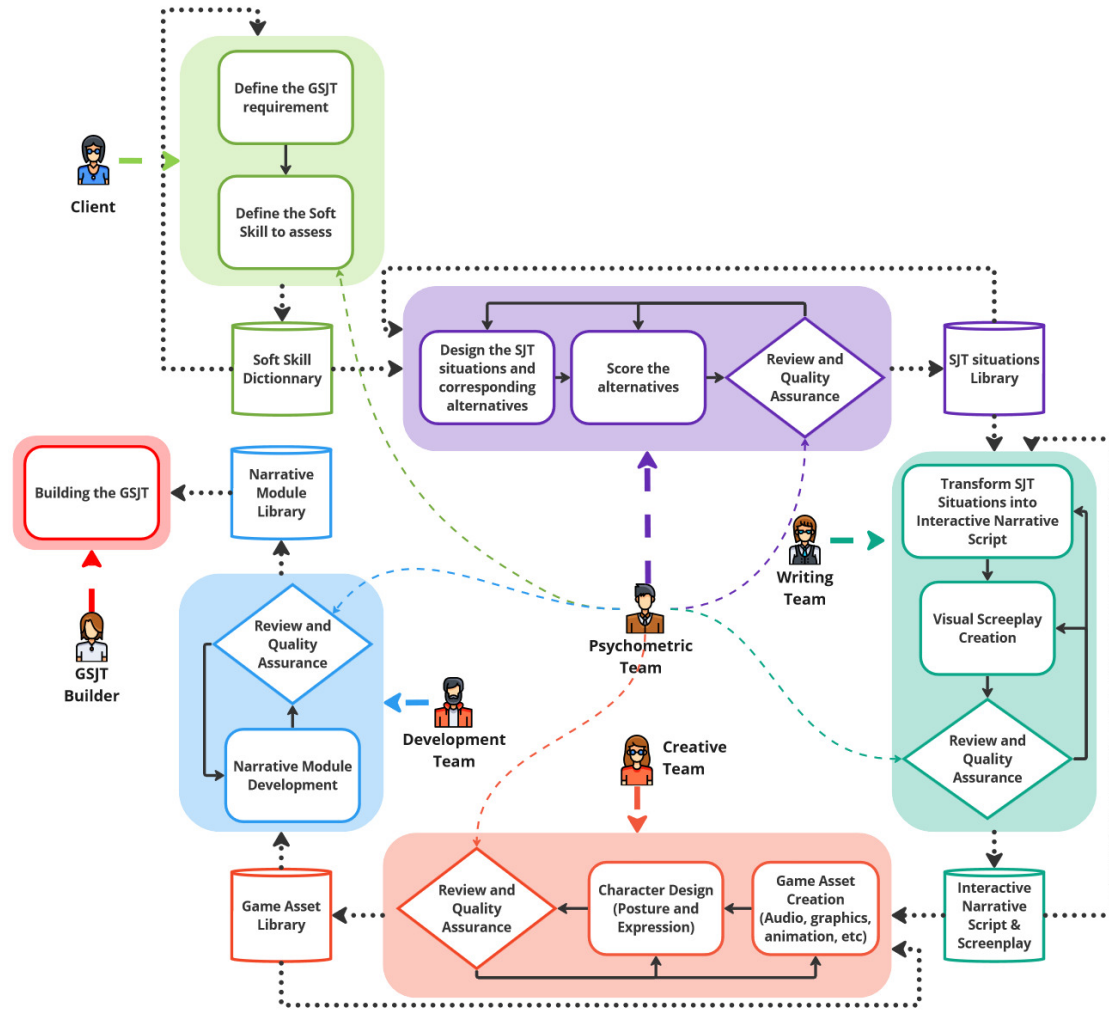


FIGURE 5.3: Gamified Situational Judgement Test Design Process in 6 Steps

We remember that the first phase involves **the client (sponsor)**, with the aid of **the Project Manager**, outlining their needs and constraints. Collaborating with **the psychometric experts**, they develop a comprehensive **dictionary of soft skills** to be assessed through the game. Afterwards, psychometric experts draw from this dictionary to identify essential soft skill aspects and **define situations and alternatives to create a traditional situational judgement test**. Game designers, including writers and scenarists, collaborate with psychometric experts to structure various **narrative modules based on the SJT situations**. This vital step involves determining which game-related scenarios align with the job-related scenarios from the SJT and facilitating player assessment. During this phase, psychometric experts must ensure the SJT maintains strong assessment potential while game designers concentrate on providing an engaging and immersive storytelling experience.

In the next phase, **the writing team** crafts the situation for every narrative module, detailing all visual, auditory, and interactive aspects. These specifications must be highly accurate to ensure understanding with **the creative team** involved in resource development and production, such as graphic designers, artists or sound designers.

Developers then use mockups and previously specified documents to create each narrative module. These prototypes undergo testing and validation by either a sample of testers or, in resource-limited companies, psychometric experts. If the test proves unsuccessful, stakeholders revisit the prior step, identify the cause, and address the issues iteratively. This process leads to an asynchronous and reliable approach to finalizing and validating each narrative module, which will be incorporated into a validated narrative module library.

Depending on the team's configuration and financial resources, individuals may assume multiple roles. In Origamix's context, for instance, a psychometric expert may also take on the writer role. Similarly, a person with video game skills might assume both a game designer and level designer roles or psychometric expert and game designer roles.

When the psychometric expert determines enough validated modules to create the entire game-based assessment, **the GSJT builders** assemble the complete GSJT like a puzzle.

The efficacy of our design methodology stems from **the precise delineation of the diverse stakeholder roles** involved in GSJT creation, as promoted by Marfisi-Schottman [45], and the adoption of a modularity approach. This latter strategy enables individuals to construct their own GSJT using the validated narrative modules library.

This second framework differs from the first one, mainly by the introduction of **a final sixth stage** called ***Building Stage*** that now may be proposed by **the GSJT builder**, who can be whoever involve in the game creation (see Figure 5.3). **This GSJT builder may modulate the GSJT as he/she pleases based on his needs and objectives.**

In addition, this novel framework transforms the production workflow revealed in Figure 4.2 and the bounds that linked the various GSJT designers' teams. Indeed, the modular approach proposed by the framework **reverses the production**

timeline as depicted in Fig. 5.4. This enables the GSJT core experts, including HR experts and writers, to create various module combinations for generating the GSJT. Indeed, these two teams must be able to fully understand each other work to keep reliability and validity in the assessment tool. In addition, the creatives and developers teams may dissociate from the GSJT core experts in the production workflow.

5.2.2 Modular Production Workflow

While the creative team will create an extensive library of creative assets required in an ING, such as soundtracks, character sprites, background, UI graphics, or videos, the development team will create an extensive library of action scripts to handle the creative assets in-game. Thus, the GSJT core experts can build in collaboration and module by module a working, valid, and reliable GSJT using a visual novel game genre.

Within the context of this novel framework, GSJT designers have the potential **to generate a multitude of GSJTs through the amalgamation of distinct narrative modules**. However, these designers must maintain an acute consciousness of the inherent narrativity of these modules to ensure the presentation of a coherent narrative.

Consequently, the next section will engage in an in-depth exploration of the proposition of a drama manager, whose purpose is **to generate GSJTs dynamically based on the narrative modules**, as mentioned earlier. This would assist GSJT designers in developing coherent interactive narrative serious games based on a semi-emergent module narrative system.

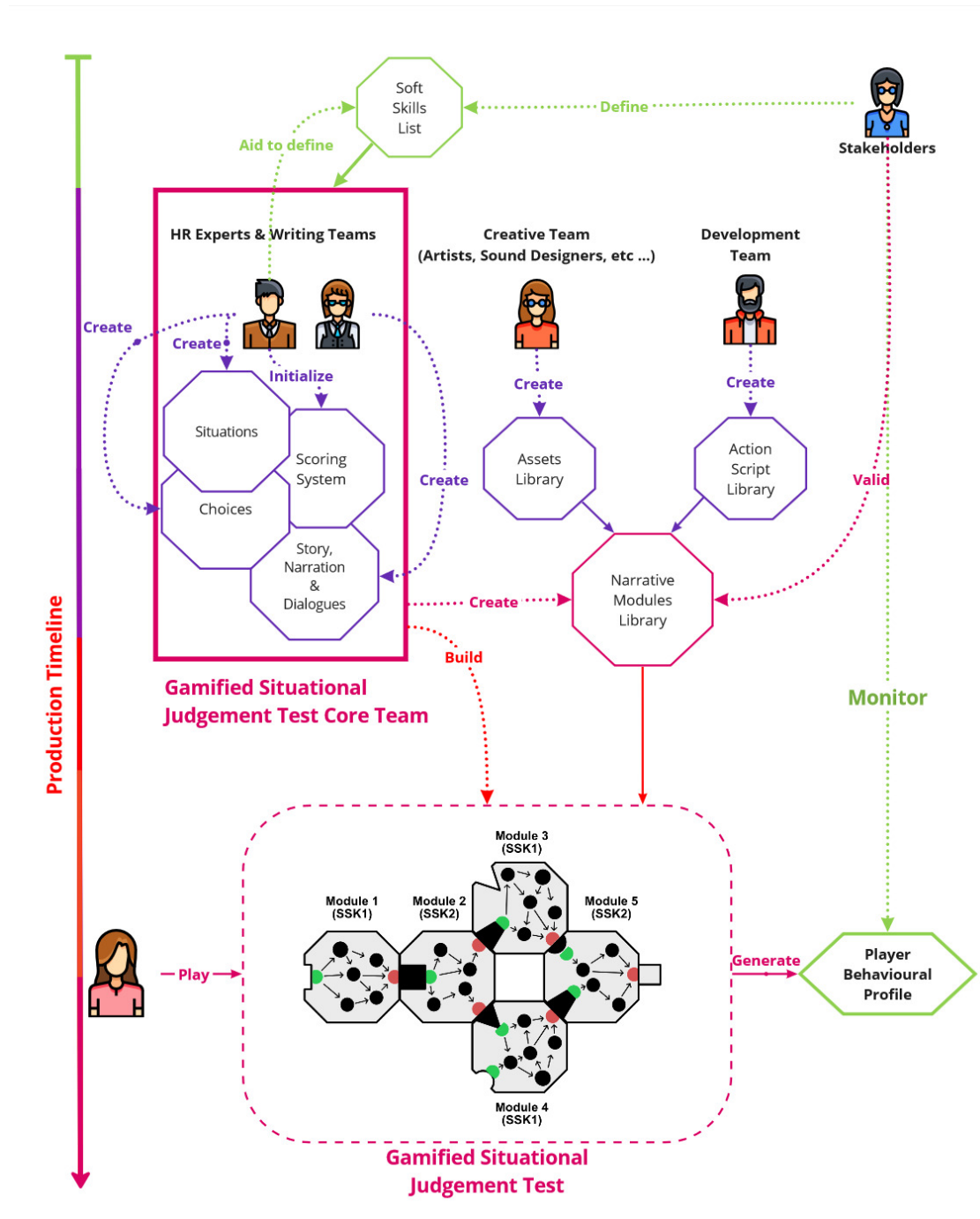


FIGURE 5.4: Modular Production Workflow

5.3 A Drama Manager to Dynamically Generate a GSJT

The implementation and use of our methodology have led us to focus on **dynamically generating gamified situational judgement tests (GSJT)** based on player choices within the game. Our primary goal is **to facilitate the efficient creation of semi-emergent modular GSJTs**. We propose that developing a drama manager that automatically selects a module after a player completes one would not only allow GSJT designers to avoid concentrating on the overall design of a GSJT, which can still be time-consuming and tedious due to the need for maintaining a coherent narrative thread with the modules but also significantly enhance player engagement.

In section 3.3.3.1, we highlighted that a drama manager is a system often used in interactive storytelling and gaming to oversee the unfolding narrative and the player's interaction. Its role is to monitor the player's actions and decisions and, based on these, dynamically adjust the game's narrative to create a satisfying and coherent story experience.

For several reasons, creating a drama manager is essential for player engagement. First, it ensures that each gameplay experience is **tailored to the individual player**, considering their choices and preferences. This personalisation fosters a sense of ownership and investment in the game, increasing the likelihood that players will remain **engaged throughout the GSJT**.

Moreover, the dynamic nature of the model **promotes replayability**, as players can explore different paths and outcomes with **each playthrough**. This variety contributes to a richer and more immersive experience, as players can discover new elements and aspects of the game that they might have missed in previous sessions.

In recruitment and selection contexts, the importance of a drama manager becomes even more critical. Providing a unique and engaging experience each time a candidate plays the GSJT may help **to prevent cheating** associated with repetitive game sessions carried out by potentially dishonest players. Candidates are less likely to share answers or strategies, as the drama manager ensures that each player's experience is unique. This ultimately may lead to a more **accurate and**

fair assessment of the candidate’s abilities, ensuring that the recruitment process remains coherent with the recruitment and selection objectives.

Therefore, we present a novel drama manager, inspired by the *”salience-based narrative”* approach employed in dialogue generation for games like *”Left4Dead”* and *”Firewatch”*, as discussed by Short [218]. Our model expands upon Short’s concept, aiming to select a narrative module from an array of options created by GSJT designers that best fits the current game state based on the player’s choices. This method remains agnostic regarding the type of information considered relevant. In our case, it may take into account in-game narrativity events, player location, or even the player’s proficiency in soft skills to suggest suitable narrative modules for evaluating players’ soft skills in line with computer adaptive testing approaches [219].

However, in this thesis, our research emphasizes narrativity due to the resource-intensive and time-consuming nature of designing and developing replayable GSJTs, which necessitates intricate scenarios, multiple storylines, and diverse challenges to maintain player interest and engagement.

Our drama manager comprises five core components: **the global game state, the narrative module selector, the narrative module library, the narrative module adaptation mechanism, and the narrative feedback system** (refer to Fig. 5.5 for details).

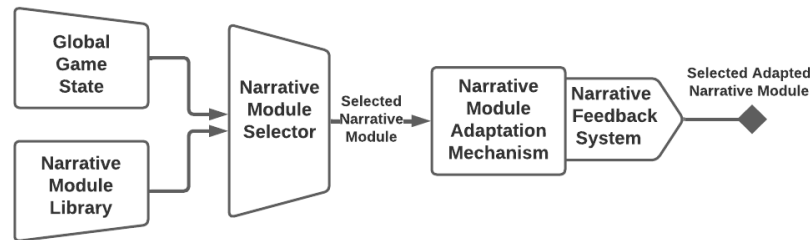


FIGURE 5.5: Dynamic Narrative Modular System

5.3.1 Global Game State

The global game state (GGS) comprehensively represents the environment essential for effective narrative module selection, adaptation, and player engagement. This state continually updates throughout the game after each player’s choice, reflecting the player’s progress and the evolving narrative context. For instance, if a

player consistently encounters communication skills narrative modules, the GGS will store this part of the information and adjust the player profile accordingly to get another narrative module that evaluates another soft skill. It also monitors the game's narrativity, such as developing a subplot or the impacts of the players on the game world. Incorporating the player's assessment profile and the narrativity inside the GGS later ensures that the model can make informed decisions and effectively respond to the dynamic nature of interactive narrative learning games.

5.3.2 Narrative Module Library

The proposed **narrative module library** is the pool of pre-defined narrative modules tailored to the client's requirements. Each module in this library is accompanied by associated **conditions and objectives**, which dictate its effectiveness and the soft skill development goals it aims to achieve. For instance, one module may concentrate on teamwork, necessitating players to collaborate with virtual characters to accomplish a task. In contrast, another module might emphasize empathy, where players must navigate a scenario involving the understanding and support of a virtual character's emotional state. In addition, each module may have several interfaces based on the soft skill to train in the module or based on the module narrativity context.

By integrating a modular design with associated conditions, objectives, and interfaces, the narrative module library enables multiple narrative pathways and the incorporation of diverse soft skill assessment scenarios. This approach promotes replayability and sustains player engagement and motivation.

5.3.3 Narrative Module Selector

The **narrative module selector** is crucial in identifying the most suitable narrative module from the library, considering the current GGS, the learner's skill development needs, and the narrative context. To achieve this, the module selector uses a heuristics approach such as trial and error (see Fig. 5.6). These heuristics enable the module selector to address the player's skill assessment needs while contributing to an engaging and motivating game experience and maintaining narrative cohesion. Based on the GGS, the narrative module selector is a dynamic interface that selects and connects modules based on tags. Guided by the

GGs, the narrative module selector functions as a dynamic interface that chooses and links modules based on tags. Each narrative module could exhibit one or multiple tags that characterize the module’s modalities. For instance, a tag may represent the location of the module, the characters’ presence within the module or conditions required based on in-game events.

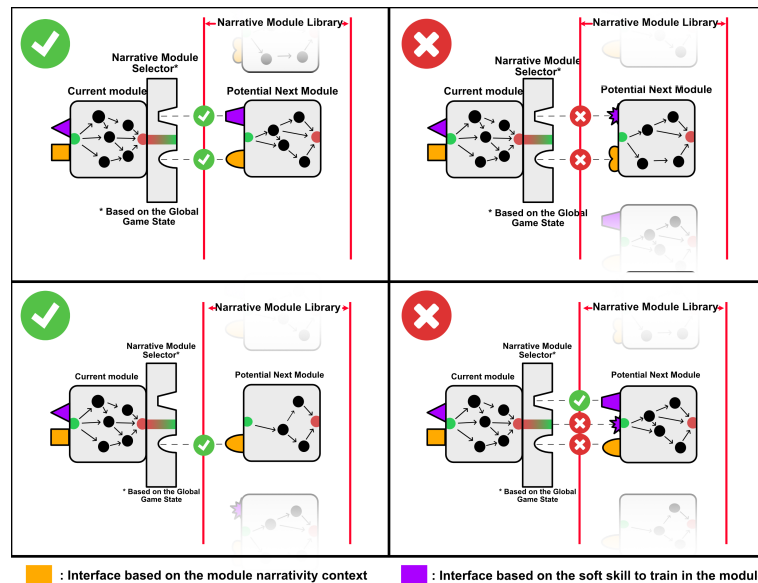


FIGURE 5.6: Four Examples of the Narrative Module Selection

For instance, consider a GSJT that evaluates leadership skills in a corporate setting within a multinational firm’s fast-paced environment. In this scenario, the narrative module selector might pinpoint a module focused on decision-making and delegation. The chosen module could integrate a situation where the player’s virtual team consists of diverse professionals with unique abilities and traits to maintain narrative consistency and merge skill development with the storyline. When faced with a tight deadline, the team requires swift resource allocation and task delegation decisions. This module challenges the learner, enabling the assessment of their leadership skills and aligning with the overarching narrative of managing a diverse team in a corporate setting. Moreover, the narrative module selector should also consider the narrative state of the world game in the narrative module selection. For example, suppose a player completes a narrative module in a certain city’s branch office. In that case, the next module should start within a regional headquarters or a different branch office in another city, ensuring everything remains cohesive in storytelling.

The narrative module selector combines the soft skill to assess with the narrative context to ensure that the selected modules are relevant, engaging, and supportive

of the assessment objectives. Considering the player's progress and the ongoing narrative, the module selector contributes a dynamic interface to a seamless and immersive experience in which skill assessment, storytelling, and imaginative settings are intricately connected. While the narrative module selector effectively identifies and presents relevant modules for the player, the narrative module adaptation mechanism and the narrative feedback system ensure these modules remain challenging, personalized, and meaningful to the player's experience.

The algorithm 1 describes how the module selector picks modules.

The algorithm we have proposed for this task organises modules into separate pools based on their start location, allowing for faster access and filtering. When the player's location changes, the algorithm retrieves the corresponding pool of modules and iterates only through that pool. It then checks if all prerequisites and conditions for each module are satisfied and whether the player has completed the required number of modules for the associated soft skill. If these criteria are met, the module is added to the list of accessible modules. By considering the player's location and the number of completed modules for each soft skill as part of the world state, the proposed algorithm identifies modules the player can access at any given time, taking advantage of the benefits of a **rule-based system**.

The rule-based system offers numerous advantages for selecting narrative modules based on player location, module prerequisites, and module-specific conditions. Some of the key benefits of this approach include:

- **Simplicity:** Rule-based systems are straightforward to implement and understand due to their reliance on simple if-then statements. This simplicity is beneficial when working with problems like module selection involving relatively simple variables and conditions.
- **Maintainability:** Rule-based systems are easier to maintain and modify than more complex algorithms. As game requirements change, developers can update the rules without restructuring the entire algorithm. This flexibility is crucial for game development, where frequent updates and modifications are often necessary.
- **Flexibility:** Rule-based systems allow for the easy addition or removal of rules as needed. This adaptability is particularly useful in game development,

Algorithm 1 Rule-Based Module Selection with Soft Skill Combination Pools

```

1: procedure SELECTMODULES(modules, worldState)
2:   softSkillPools  $\leftarrow$  ORGANIZEMODULESBYSOFTSKILLCOMBINATION(modules)
3:   accessibleModules  $\leftarrow$  empty list
4:   for each softSkillPool in softSkillPools do
5:     if not ALLSKILLSBELOWN(softSkillPool, worldState, N) then
6:       continue
7:     end if
8:     for each module in softSkillPool do
9:       if not LOCATIONMATCHES(module, worldState) then
10:        continue
11:      end if
12:      if not PREREQUISITESATISFIED(module, worldState) then
13:        continue
14:      end if
15:      if not CONDITIONSMET(module, worldState) then
16:        continue
17:      end if
18:      accessibleModules.append(module)
19:    end for
20:  end for
21:  return accessibleModules
22: end procedure

23: function ORGANIZEMODULESBYSOFTSKILLCOMBINATION(modules)
24:   softSkillPools  $\leftarrow$  empty dictionary
25:   for each module in modules do
26:     softSkillKey  $\leftarrow$  SORTANDHASH(module.softSkills)
27:     if softSkillKey not in softSkillPools then
28:       softSkillPools[softSkillKey]  $\leftarrow$  empty list
29:     end if
30:     softSkillPools[softSkillKey].append(module)
31:   end for
32:   return softSkillPools
33: end function

34: function ALLSKILLSBELOWN(softSkillPool, worldState, N)
35:   for each softSkill in softSkillPool do
36:     if worldState.completedModulesBySkill[softSkill]  $\geq N$  then
37:       return false
38:     end if
39:   end for
40:   return true
41: end function

42: function LOCATIONMATCHES(module, worldState)
43:   if module.startLocation = worldState.playerLocation then
44:     return true
45:   else
46:     return false
47:   end if
48: end function

```

where module selection logic may need to be modified based on user feedback or design changes.

- **Performance:** Given the simplicity of the module selection problem, a rule-based system is likely to provide adequate performance without the need for more complex algorithms. By efficiently evaluating conditions and prerequisites, rule-based systems can quickly determine which modules the player can access at any given time.

5.3.4 Narrative Module Adaptation Mechanism and Narrative Feedback System

The narrative **module adaptation mechanism** also uses the GGS to continuously adapt and improve the adequacy of narrative modules, including the dialogues and narrativity within each module. For example, suppose a player has shown a strong understanding of negotiation techniques. In that case, the adaptation mechanism might modify the dialogues and narrative elements within a negotiation-focused module to present more challenging situations, pushing the assessed candidate to deploy and show their skills further. Similarly, if a choice in the player's path impacts a character's psychological state, it will adjust the virtual character's dialogue accordingly. Adjusting the dialogues and narrativity within each module based on the GGS ensures that the narrative content remains contextually relevant, engaging, and motivating for the player.

On the other hand, the **Narrative Feedback System (NFS)** provides direct and delayed narrative feedback to players through dialogues and scenarios based on the GGS. For example, after a player successfully resolves a conflict, they might receive immediate feedback through dialogue with a virtual character who thanks them for their help. In a later scenario, the player may encounter a situation that references the previously resolved conflict, providing delayed feedback that reinforces the immersive experience. By integrating dynamic feedback and make-believe elements, the NFS contributes to the overall effectiveness of the extended method, ensuring that players remain engaged, motivated, and immersed.

5.3.5 Conclusion

In conclusion, our modular drama manager for interactive narrative learning games comprises five main components: **the global game state, the narrative module selector, the narrative module library, the narrative module adaptation mechanism, and the narrative feedback system**. By incorporating these elements and continuously updating the GGS, our model ensures that players remain engaged, motivated, and immersed, ultimately leading to more effective soft skill evaluation and behavioural skill recognition. Likewise, such drama managers enable the creation of a unique experience for the game sessions and may help the GSJT designer create an ING for soft skills recognition. Therefore, the following section will explore using the framework and the manager to generate a novel GSJT called *Shogureign*.

5.4 Experiment 2: A second GSJT, Experimental Design and Objectives

This section will present an experimental study to investigate **the effectiveness of the dynamic narrative modular model** presented in Chapter 5.3. The objectives are to explore the influence of a dynamic modular design for GSJT, enabling the automatic generation of contextually relevant scenarios for a player, thereby facilitating the development of GSJT and enhancing the potential of replayability for recruitment and selection purposes.

To achieve these goals, the study will be designed to address the following research question:

- **A dynamic GSJT generation approach may also propose an efficient way to enable the replayability of GSJT.**

In order to address this hypothesis, the experimental study will involve a comparison between a GSJT utilizing the dynamic approach for combining narrative modules, which were created using **the six-step methodology**, and its corresponding traditional survey. Participants will engage in both assessment methods, completing a series of scenarios designed to evaluate six specific soft skills based on **the**

HEXACO model [220]. Data concerning player engagement, immersion, and assessment outcomes will be gathered and analyzed throughout the experiment.

By scrutinizing these factors, the goal is to obtain a more comprehensive understanding of the advantages and potential limitations of employing the dynamic narrative modular approach in creating GSJTs. This insight can then be harnessed to fine-tune and optimise the model, ultimately aiding in developing more efficient, captivating, and personalised gamified situational judgement tests. Additionally, this study provides preliminary evidence regarding the potential for **replayability in such gamified assessment methods**.

5.4.1 The Experimental Gamified Situational Judgement Test and SJT

The original design of the experimental GSJT aimed to evaluate all **twenty-four HEXACO soft skills**, also referred to as **HEXACO sub-facets**. This was accomplished by designing **87 narrative modules**, enabling the creation of diverse GSJT combinations. However, several assessment modules created needed to be validated by the stakeholders and subject matter experts. Therefore the validated experimental GSJT can assess **six soft skills** based on HEXACO through **42 narrative modules**: (1) *Aesthetic Appreciation*, (2) *Innovation*, (3) *Inquisitiveness*, (4) *Prudence*, (5) *Liveliness*, and (6) *Anxiety control*. The GSJT designers team comprised **three writers, one graphical designer and one developer**. The GSJT designers based the gamified situations and choices on an existing situational judgement test that evaluates a sole, soft skill per situation following the HEXACO model [185]. **26 of the 42 narrative modules are assessment modules** (see Figure 5.7). The other modules are used in the GSJT for **stealth assessment and entertainment purposes**. The length of each module may vary depending on the narrative and interactive content.

In the experimental GSJT, the player assumes the role of a Sino-Japanese monarch and is tasked with ensuring the prosperity of his/her kingdom through effective management.

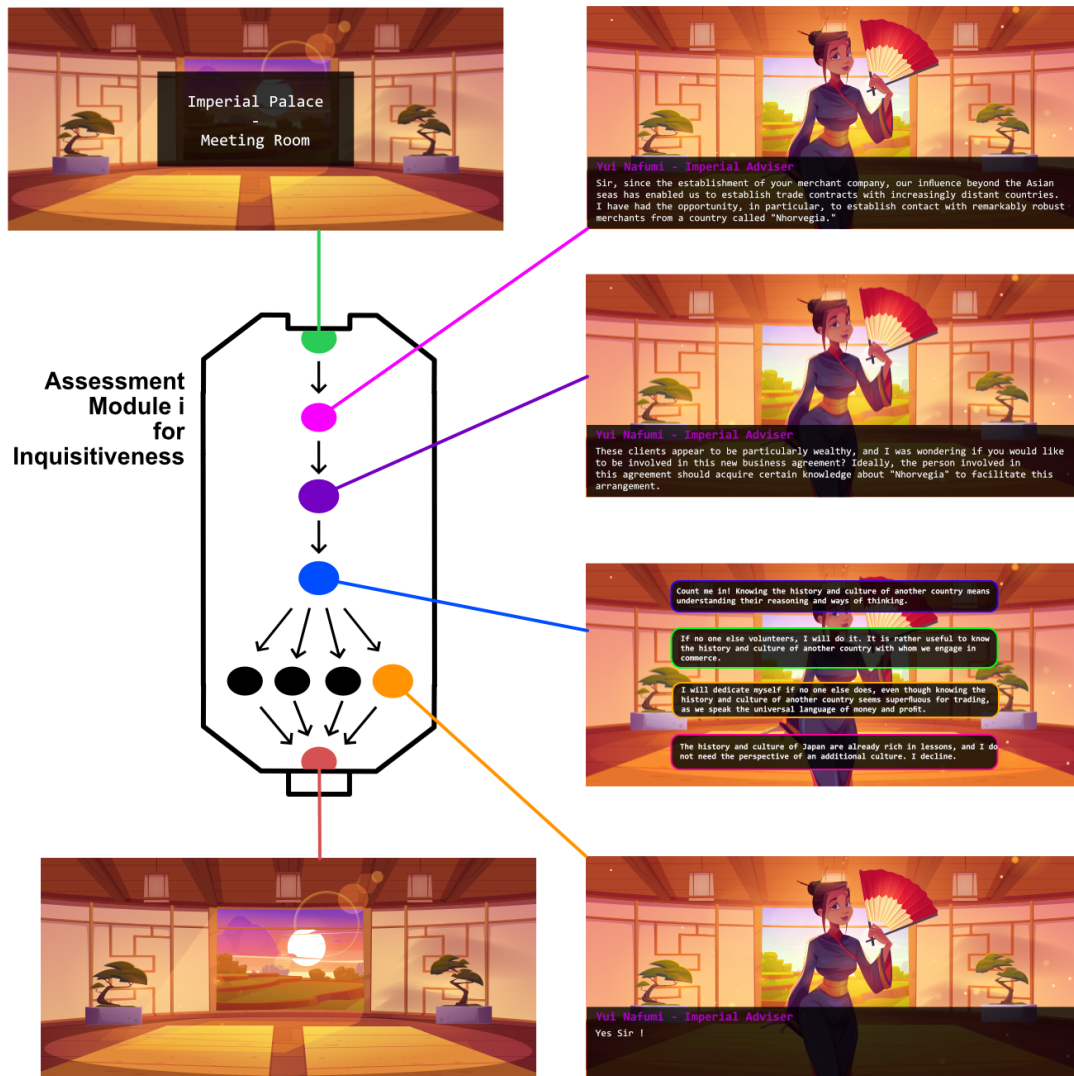


FIGURE 5.7: Example of an Assessment Module which assess "Inquisitiveness" soft skill in the Experimental GSJT

Note: Character sprite and background are not from the actual game and are adapted from free-to-use assets available on Freepik and created by 'upklyak' sur Freepik for illustration purposes.

5.4.2 Sample and Procedures

Test-takers/players were invited online to complete the traditional 60-item HEXACO survey [220]. Participants were from five countries, Mali, Niger, France, Belgium, and Tunisia and represented 12 different socio-professional categories and 21 distinct industry sectors. Eighty-five test-takers completed the survey, while 44 played the GSJT with a one-week gap between the tests to minimize potential biases.

The average time to complete the survey was approximately **14 minutes**, while the GSJT took about **25 minutes**. Notably, each player experienced **a unique path** within the GSJT. After the game, **13 players were interviewed** for qualitative feedback on their experiences.

Ultimately, **two types of behavioural profiles were generated** for each player, representing the combination of the six assessed soft skills: (1) **The HEXACO survey behavioural profile**, which used 2 or 3 items to evaluate each soft skill, and (2) **The GSJT behavioural profile**, which assessed each soft skill using **four items/situations (modules)**.

This diverse and international participant pool provided valuable insights into the effectiveness and adaptability of the GSJT across various backgrounds and industries. The study also offered an opportunity to compare the assessment results from the traditional HEXACO survey and the GSJT, shedding light on the potential benefits and challenges of implementing a gamified approach to soft skill evaluation.

Finally, **seven players were interviewed** to collect insight into their gamified assessment experience.

5.4.3 Results and Discussion

The hypothesis predicts that a dynamic GSJT generation approach may also propose an efficient way to **enable the replayability of GSJT for recruitment and selection purposes**. The hypothesis suggests that we keep coherence in the assessment objectives of the serious game. Thus, we first perform **a criterion validity analysis for the GSJT using the Pearson correlation**. We calculate the correlation between **the GSJT results and the results of the HEXACO 60-items as the criterion measurement**. Conversely, we perform **an internal consistency using Cronbach's alpha** for both measurement tests, the HEXACO 60-items and the GSJT.

TABLE 5.1: Intercorrelations for all variables

Assessment Method	Soft Skill	GSJT						HEXACO Survey					
		1	2	3	4	5	6	1	2	3	4	5	6
GSJT	1	1											
	2	0.28	1										
	3	0.34	0.01	1									
	4	0.29	0.1	-0.01	1								
	5	0.48	-0.04	-0.01	0.34	1							
	6	0.47	0.3	0.3	0.2	0.06	1						
HEXACO Survey	1	0.48	0.43	0.08	0.4	0.08	0.26	1					
	2	0.4	0.29	0.45	0	-0.19	0.19	0.44	1				
	3	0.13	0.36	0.11	-0.08	-0.25	0.05	0.59	0.7	1			
	4	0.23	0.07	0.11	0.2	0.13	0.07	-0.09	0.05	-0.01	1		
	5	0.35	-0.06	0.08	0.07	0.2	-0.11	-0.04	0.07	-0.21	0.27	1	
	6	0.06	-0.07	0.11	0.26	-0.2	0.12	0.28	0.05	0.12	-0.1	-0.13	1

Note. N = 44.

The boldface indicates the Pearson correlation measures between the corresponding soft skills of the two assessment methods.

5.4.4 Tests of the hypothesis

In Table 5.1, we observed a **positive correlation coefficient** between the two assessment methods for each corresponding soft skill, indicating a positive relationship and supporting our hypothesis that our semi-emergent modular GSJT approach may generate various coherent GSJT with the objectives of assessment of soft skills. However, the correlation coefficient ranges from 0.11 to 0.48, suggesting a **weak to moderate relationship** between the two assessment methods (GSJT and SJT). However, the internal consistency analysis may explain such moderate results.

According to academic literature, a value of **0.7 or higher** is considered acceptable to indicate items' internal consistency in measuring the same underlying construct of soft skills. While poor internal consistency for five of the six assessed soft skills may suggest that the HEXACO 60-item survey is not valid in our experimental context (see Table 5.2), it is crucial to consider potential sources of bias and the need for further replication studies with extensive samples.

Conversely, we observed **high internal consistency** between the items assessing the same soft skills in the modular GSJT. Cronbach's alpha values ranged from **0.65 to 0.87** for the six assessed soft skills. Our study's results support that implementing a semi-emergent modular GSJT structure can effectively enhance reuse potential across various scenarios while maintaining a balance between engagement and the assessment purpose of GSJT.

TABLE 5.2: Means, standard deviations, and internal consistency reliability

Assessment Method	Soft skill	<i>M</i>	<i>SD</i>	Cronbach's α
GSJT	1	3,2	0,37	0,65
GSJT	2	3,47	0,5	0,75
GSJT	3	2,9	0,71	0,81
GSJT	4	2,95	0,86	0,87
GSJT	5	3,11	0,6	0,76
GSJT	6	2,33	0,81	0,81
HEXACO	1	3,92	1,08	0,45
HEXACO	2	3,79	0,77	0,46
HEXACO	3	3,64	1,12	0,65
HEXACO	4	3,63	1,48	0,47
HEXACO	5	3,83	1	0,8
HEXACO	6	3,52	1,44	0,81

Note. N = 44

The boldface indicates significant internal consistency reliability.

The GSJT soft skills were measured on a scale from 1 to 4.

The HEXACO survey soft skills were measured on a scale from 1 to 5.

Consequently, these results indicate that the semi-emergent narrative modular model can create various GSJTs while maintaining coherence between assessment objectives and in-game narrativity, providing preliminary results for a **replayability potential using narrative modules for soft skill recognition**. Additionally, GSJT designers can use fewer items per soft skill in the GSJT and replace the others with entertaining situations to enhance engagement and motivation in-game. Indeed, players' interviews revealed that they enjoyed playing the game and were more willing to participate in such assessments than traditional ones. However, they could sometimes sense they were being assessed in specific situations. Following the internal consistency, proposing fewer assessment modules that assess the same construct enables more entertainment modules, enhancing validity through stealth assessment and reducing stress anxiety.

Furthermore, interviews revealed that players enjoyed the narrativity **without noticing weak connections between situations**. This outcome indicates that despite having **three different writers produce narrative modules**, a modular approach enhances productivity while maintaining overall storytelling and narrativity.

These results suggest that a modular design may improve the GSJT production

workflow. Although no HR experts were involved in the GSJT production, **a high-reliability value** implies that even **non-experts can create a GSJT** easily if they have a traditional SJT created and initialized by a psychometric expert to convert.

In summary, these findings provide evidence that companies with limited resources and subject matter expertise may create engaging recruitment and selection tools using a modular GSJT design.

5.4.5 Conclusion

Our analysis shows that the behavioural assessment conducted through the generated modular GSJT is coherent with the traditional HEXACO 60-item survey [220]. Furthermore, our experimental results demonstrate that the modular approach streamlines the production workflow, efficiently facilitating the creation of diverse and unique GSJTs. This experimentation and this chapter's proposals have been the subject of a peer-reviewed scientific paper [221].

Nonetheless, several limitations must be acknowledged in this study:

- **The results need validation using a more extensive and diverse dataset to strengthen the validity and reliability of our findings.**
- **While the behavioural assessment demonstrates positive criterion validity, additional research is required to determine predictive validity and ensure the complete effectiveness of this approach for selection and recruitment applications.**
- **The study reveals that although players enjoyed the game, they could sense assessment in certain situations, which may introduce bias into the behavioural assessment outcomes.**

In future research, addressing these limitations and exploring additional factors that can further enhance the modular GSJT design will be essential. For example, investigating **the impact of stealth assessment techniques or incorporating adaptive difficulty levels** could improve player engagement while reducing the perception of being assessed. Moreover, examining the effectiveness of this approach across various industries and job roles will provide valuable insights

into its **generalisability and practical applicability**. By building on the current study's findings, researchers can continue to refine and **optimise the semi-emergent modular GSJT generation approach**, ultimately contributing to a more engaging, valid, and reliable tool for assessing soft skills across various contexts and user groups.

Chapter Overview

In alignment with *research objective A*, which is to develop a game-based behavioural assessment method, this chapter proposes a **Semi-Emergent Modular Gamified Situational Judgement Test** using **narrative modules and interfaces**.

- **Narrative modules** are fragments that form a cohesive GSJT storyline when combined.
- **Interfaces** set the interaction rules between different modules.

In the second part, we outlined a framework for a semi-emergent modular GSJT model and a workflow for efficient GSJT design, aiming to present preliminary results for *research objective B*. This proposal, based on the chapter 4 framework, introduces the new **GSJT Building** stage and **GSJT Builder** role, facilitating the final GSJT's creation.

In the third section, we explored a **Dynamic Narrative Modular System** with five core elements, aiming for an immersive, unique assessment experience.

- **The global game state & the narrative module selector.**
- **The narrative module library & The narrative module adaptation mechanism.**
- **The narrative feedback system.**

Finally, an experiment was conducted to test **the second framework** described in the thesis, focusing on its design, execution, and resulting data. We examined the experimental GSJT's **effectiveness and replayability** using the proposed drama manager alongside a traditional SJT and outlined the sample and procedures used in the experiment to gather findings related to our research objectives.

Chapter 6

Proposing a Computational Psychometric Scoring System to Help GSJT's Psychometrics Teams

In order to further elaborate on the frameworks proposed in chapters 4 and 5, it is essential to note the crucial role each stakeholder plays in the recruitment and selection process. The defined roles and assigned tasks guide each stakeholder to ensure a structured and efficient workflow. However, as highlighted in the Introduction 1, the complexities of the recruitment and selection field may result in some **stakeholders assuming multiple roles or performing tasks outside their domains of expertise**.

These complexities must be considered to maintain the reliability and validity of the final assessment. The stakeholders in the psychometrics and writing teams hold a crucial responsibility in this regard, as they are responsible for ensuring that the assessment accurately reflects the abilities and qualities of the evaluated candidates. To support these endeavours, the proposed approach for psychometrics and writing team aims to address any potential challenges and enhance their respective processes.

Therefore, by considering the complexities of the context and providing support for the psychometrics and writing teams, the two following chapters aim to propose approaches to assist in the development of the GSJT.

We highlighted earlier that the psychometric team plays a vital role in this process, as they are responsible for creating and employing a scoring method that accurately reflects the abilities and qualities of the evaluated candidates. In this current chapter, we will develop **an approach to help psychometrics experts propose scoring matrices** for the GSJT as scoring GSJT scenarios is critical in determining the reliability of the assessment [222].

6.1 Scoring a Situational Judgment Tests

In this section, we will explore **the various scoring methods of situational judgment tests (SJT)** in the context of soft skill recognition. Scoring methods play a crucial role in accurately identifying and assessing an individual's soft skills, thus contributing to these tests' overall validity and reliability. By examining different scoring methods, we aim to highlight their significance in the assessment process and provide a comprehensive understanding of their strengths and weaknesses.

We will begin by providing an overview of the most common scoring methods used in SJTs, such as **consensus scoring, expert-based scoring, and rational scoring, among others**. We will discuss each method's underlying principles, applications, and potential limitations.

Following the examination of the different scoring methods, we will present a critical analysis of their strengths and weaknesses. This analysis will help us identify key characteristics essential for accurately assessing soft skills in our context. By understanding these characteristics, we aim to integrate them into our research methodology, ultimately enhancing our study's effectiveness in soft skill recognition through situational judgment tests.

The analysis of the strengths and weaknesses of these scoring methods will allow us to identify a list of characteristics that we consider essential in the proposition of evidence for the identified hypothesis:

- **The application of computational psychometrics may assist psychometric experts in developing a relevant in-game soft skill measurement system.**

We will explore the following seven SJT-specific scoring methods: Consensus Scoring, Expert-based Scoring, Empirical Key Scoring, Rational Key Scoring, Subgroup Normative Scoring, Dimension-Based Scoring, and Weighted Scoring.

6.1.1 Scoring Methods: Modalities, Advantages and Limitations

- Consensus Scoring:** Consensus scoring compares the test-takers' responses to the most common responses given by a reference group, such as experts or peers [223]. The underlying assumption of this method is that the majority response within the reference group is the most appropriate or accurate response to a given situation. To implement consensus scoring, researchers first administer the SJT to a representative reference group of subject matter experts, successful job incumbents, or a diverse sample of individuals from the target population. The most common response for each item is identified and designated as the *"correct" or "optimal" answer*. Test-takers' scores are then calculated by comparing their responses to these consensus answers, with higher scores indicating greater alignment with the majority opinion of the reference group. One of the main advantages of consensus scoring is its relative ease of implementation, as it does not require extensive input from experts, nor does it rely on complex statistical procedures. Additionally, consensus scoring can be instrumental in situations where there is no clear-cut correct answer, as it reflects the collective wisdom of the reference group. This method can also effectively reduce the impact of individual biases, as it aggregates the opinions of multiple individuals.
- Expert-Based Scoring:** Expert-based scoring needs subject matter experts (SMEs) to rate the effectiveness of each possible response option [184, 224]. Experts in this approach are typically chosen based on their extensive knowledge and experience in the assessed domain. The experts review each SJT item and evaluate the appropriateness of each response option, often assigning a numerical rating to indicate the effectiveness of the response in addressing the specific situation. The average of the expert ratings for each response option is then used as a benchmark to score the test-takers responses. One of the primary advantages of expert-based scoring is its potential for high content validity, as it incorporates the knowledge and experience

of professionals well-versed in the domain being assessed [225]. This method can provide an accurate and reliable assessment of test-takers' abilities by reflecting the judgments of those who have a deep understanding of the competencies required for success in the target job or role. However, there are several potential drawbacks to expert-based scoring. First, the process can be time-consuming and resource-intensive, as it requires the involvement of multiple experts, whose time may be limited and costly. Additionally, the process may be susceptible to subjectivity and bias, as experts might have differing opinions on the best response or be influenced by personal beliefs and experiences.

- **Empirical Key Scoring:** Empirical key scoring derives the correct answer for each test item based on the responses of a large sample of individuals who are successful in the target job or role [225]. This approach assumes that individuals who demonstrate success in the assessed domain most frequently choose the most effective response. Researchers first administer the SJT to a representative sample of successful job incumbents or high-performing individuals within the target population to implement empirical key scoring. The most common response for each item is identified and designated as the *correct* or *optimal* answer. One of the primary advantages of empirical key scoring is that it is grounded in the actual performance and choices of successful individuals, which can enhance the criterion-related validity of the SJT scores [194]. This method can also be relatively easy to implement, as it does not require extensive input from subject matter experts or rely on complex statistical procedures.
- **Rational key scoring:** In rational key scoring, the correct answer for each test item is determined through a logical, deductive process, such as expert consensus or a priori reasoning [223, 225]. This method is based on the premise that a theoretically sound and coherent set of response options can be derived from the principles and theories underlying the construct being assessed. For example, if an SJT measures ethical decision-making, the correct answers may be determined by referring to established ethical principles and guidelines. Rational key scoring can be advantageous when empirical data is limited or difficult to obtain, as it relies on the logical coherence of the response options rather than empirical evidence. This method is advantageous when the assessment measures abstract constructs, such as moral

reasoning or ethical decision-making, where obtaining empirical data can be challenging [31].

- **Subgroup Normative Scoring:** Subgroup normative scoring involves comparing test-takers scores to the average scores of different subgroups based on age, gender, ethnicity, or educational background [226]. This method allows for more tailored and fair comparisons by accounting for potential differences in test performance that these factors may influence. By comparing test-takers' scores within their respective subgroups, this scoring method can help mitigate the effects of potential biases and enhance the fairness of the assessment. One of the primary advantages of subgroup normative scoring is that it promotes greater equity and fairness in the assessment process, mainly when dealing with diverse populations. This method can help identify and control for potential sources of bias in the test content or administration, ensuring that the scores accurately reflect the abilities and skills of the test-takers.
- **Dimension-based scoring:** Dimension-based scoring involves scoring test items based on the specific dimensions of the construct being assessed, such as communication skills, decision-making, or leadership abilities [225]. This method allows for a more detailed evaluation of test-takers' skills and competencies by examining their performance on each dimension separately. By providing insights into the strengths and weaknesses of the test-takers across different dimensions, this method can facilitate targeted feedback and tailored interventions for improvement. One of the main advantages of dimension-based scoring is its ability to provide a granular assessment of test-takers abilities across multiple dimensions of the measured construct. This can be particularly useful in situations where it is essential to identify specific areas of strength or weakness, such as in employee selection and recruitment processes.
- **Weighted scoring:** In weighted scoring, different response options are assigned different weights based on their relevance or importance to the construct being assessed [225]. This method allows for a more nuanced evaluation of test-takers performance by considering the relative importance of each response option within a given test item. Weighted scoring can be instrumental in situations where some response options may be more desirable or practical than others, but all options are strictly correct and appropriate.

Despite their advantages, these scoring techniques possess inherent limitations, posing challenges in test validity, reliability, and generalizability. One such limitation is the assumption that the majority response is always the best or most appropriate, which is inherent to consensus scoring, empirical key scoring, and dimension-based scoring methods [227]. In these methods, the majority response of a reference group, successful individuals, or dimensions, respectively, are designated as the correct or optimal answer. This assumption might not always hold true as the majority opinion could be influenced by biases, misconceptions, or groupthink, leading to suboptimal response choices. Additionally, these methods' effectiveness is **contingent on the representativeness and quality of the reference group or sample**. If these groups or samples inadequately represent the target population or lack the necessary knowledge and expertise, the validity and reliability of the scores might be compromised.

Expert-based and rational key scoring methods have a common limitation regarding their susceptibility to subjectivity and bias [184]. As these methods rely on the judgments of subject matter experts or the assumptions of a specific theory, **biases and personal experiences** could influence the scoring decisions. This might lead to inconsistencies in scoring and interpretation, affecting the reliability and fairness of these methods. This issue could be particularly pronounced in situations where only a few SMEs are involved in the scoring process, such as in limited-resources companies.

Subgroup normative scoring may present a unique set of limitations related to the criteria used to define subgroups and the size and diversity of the sample [228]. If subgroup norms are based **on small or unrepresentative samples**, new biases or inaccuracies may be introduced. Additionally, establishing reliable and valid subgroup norms can be challenging due to the need for larger and more diverse samples.

For dimension-based scoring, a major limitation is the requirement for **a clear and well-defined conceptualisation of the dimensions** of the assessed construct [225]. Developing items that accurately measure each dimension can be challenging, and issues related to the discriminant validity of the dimensions might arise if the items are not meticulously designed and validated.

Finally, weighted scoring has a unique challenge related to **the assignment of**

weights to different response options [225]. A clear and well-justified rationale is needed for weight assignment, which might require substantial input from subject matter experts or empirical data. This method may also introduce greater complexity in scoring and interpretation, which could pose a concern for test designers and administrators.

6.1.2 Conclusion

In conclusion, this section comprehensively examined various scoring methods employed in situational judgment tests (SJTs) for soft skill recognition, including consensus scoring, expert-based scoring, empirical key scoring, rational key scoring, subgroup normative scoring, dimension-based scoring, and weighted scoring. The analysis has demonstrated that each scoring method **offers unique advantages and challenges**. These methods may play a vital role in ensuring the validity and reliability of SJTs in assessing an individual's soft skills. By critically analysing the strengths and weaknesses of each method, we have identified key characteristics essential for the accurate assessment of soft skills in our context, which will be integrated into our research methodology to enhance the effectiveness of our study and subsequent proposal.

6.2 A Computational Psychometric Scoring Approach

In a GSJT, the alternatives serve as **indicators that monitor the player's decision-making process**, ultimately enabling **the definition of the player's behavioural profile via scoring matrices** upon game completion. However, as previously noted, the various gamification processes employed in SJTs may impede coherence in the scoring matrices between the SJT scoring process and the creation of narrative modules. Furthermore, the verification process of each scoring matrix during the narrative module development stage may prove laborious, time-consuming, and economically inefficient, particularly for organisations with limited resources or those relying on external entities for the review process. In light of these challenges, we propose **a computational psychometric scoring approach**. Based on the proportion consensus and weighted scoring methods,

as elaborated in the previous section, the computational psychometric scoring approach aims **to assist psychometric experts in refining the scoring matrices** of each narrative situation during the fifth stage of our frameworks. Such approach may streamline the review and quality assurance process compared to the exclusive use of a human scoring approach to generate a behavioural profile.

In this section, we will first explain **how a behavioural profile in a GSJT is generated using a scoring system based on response vectors (RV)** defined by the psychometrics experts and **the player path**. Then, we will explain our contribution to enhance this process through a computational psychometric approach.

6.2.1 Scoring Matrices, Behavioural Profile and Scoring System

Our thesis context suggests that the dynamic nature of player choices in the GSJT offers a unique approach to evaluating player behavioural profiles, as it allows for the continuous assessment of player choices throughout the game. As players navigate various scenarios and make decisions, their behavioural profile fluctuates, capturing the dynamic nature of human behaviour.

A behavioural profile can be understood as **a multidimensional representation of an individual's tendencies, habits, and preferences in different contexts**. It encompasses various aspects of human behaviour and personality encompassed in the Introduction 1 under **the terms soft skill (SSk)**. In the given context, a player's behavioural profile is precisely characterised by the final score in each soft skill upon completing the GSJT, as illustrated in Figure 6.1.

The value of the score may be varied following a rating scale provided by the psychometric team in the second phase of our previous methodology (see Table 6.1).

This same rating scale enables the psychometric team to define the scoring matrices for each narrative module. Indeed, through the narrative modules, players are presented with various situations that require them to make choices that directly impact their behavioural profile and draw their respective game paths. Each alternative and the player path within a narrative module may be represented by a decision graph as depicted in Figure 6.2.

	SSK1	SSK2	SSK3	SSK4
Player 1	2	1	4	4
Player 2	5	2	2	5
Player 3	2	4	2	1

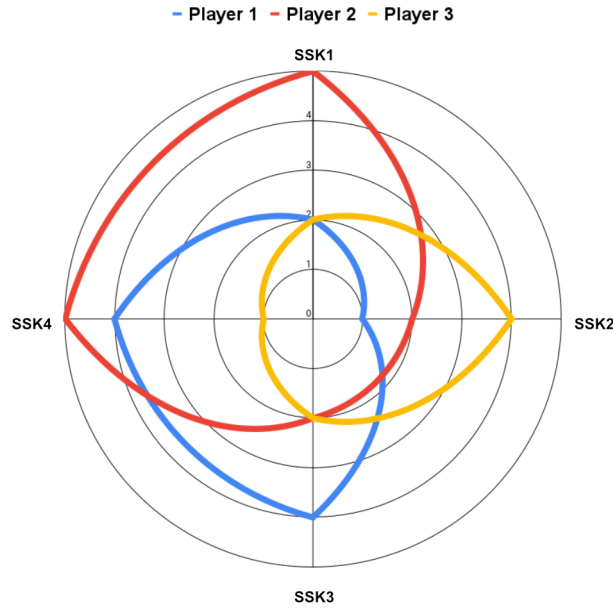


FIGURE 6.1: Example of Three Behavioural Profiles

TABLE 6.1: Example of a Rating Scale

Score	Description
5	Extensive experience in the soft skill area
4	Good experience in the soft skill area
3	Some experience in the soft skill area
2	Little experience in the soft skill area
1	No experience in the soft skill area

In the graph, every node symbolises **a distinct alternative** that the player may potentially select within a narrative module, as depicted in Figure 6.2. These scenario alternatives are expressed using an n -dimensional vector, referred to as **the response vector** RV . The response vector's components are representative of the significance of the SSks, which are linked to the respective scenario alternatives being considered. Therefore, **a narrative module scoring matrix** $NMSM$ contain the RV of each situation alternative (eq. 6.1).

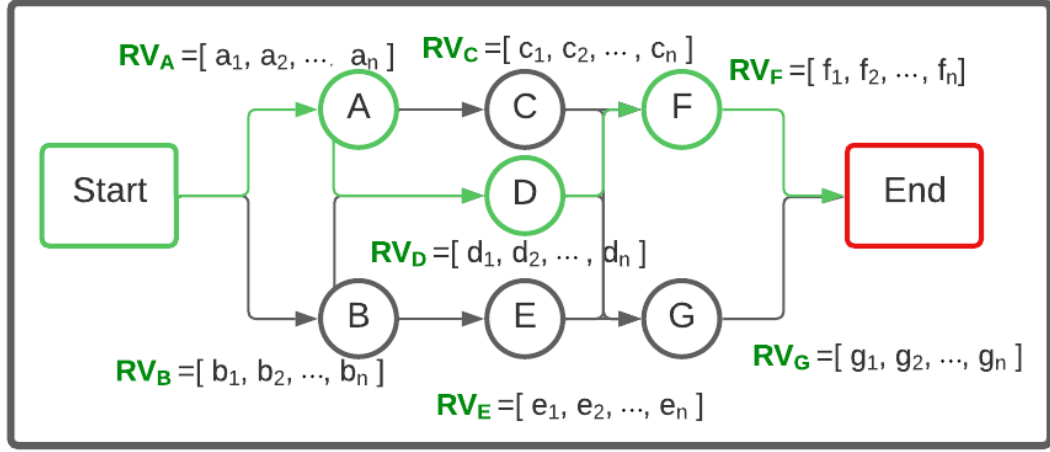
Narrative Module

FIGURE 6.2: An example of a Narrative Module Decision Graph

$$NMSM = \begin{bmatrix} RV_A \\ RV_B \\ RV_C \\ RV_D \\ RV_E \\ RV_F \\ RV_G \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \\ c_1 & c_2 & \dots & c_n \\ d_1 & d_2 & \dots & d_n \\ e_1 & e_2 & \dots & e_n \\ f_1 & f_2 & \dots & f_n \\ g_1 & g_2 & \dots & g_n \end{bmatrix} \quad (6.1)$$

Finally, using the scoring matrices with their respective response vectors combined with the player path, we can define a respondent matrix R through a scoring system. In **the matrix** R , each row corresponds to **a response vector** RV that is influenced by the decisions made by the player throughout the game. For example, as shown in Figure 6.2, the player is initially presented with **two options (A and B)**. If the player opts for **alternative A**, RV_A is incorporated into their chosen path. Upon making their final choice, the system computes the average of each column in **the matrix** R to derive the **player's behavioural profiles BP**. Consequently, the behavioural scoring mechanism represents a player's behavioural profile as assessed by the game using **an n -dimensional vector**, where n **indicates the total number of soft skills evaluated**.

The following equations represent, respectively, the path of a player P (eq. 6.2), the corresponding respondent matrix R (eq. 6.3) and the behavioural profile (eq. 6.4) generated from the GSJT example depicted in Fig. 6.2.

$$P = \begin{bmatrix} A & D & F \end{bmatrix} \quad (6.2)$$

$$R = \begin{bmatrix} RV_A \\ RV_D \\ RV_F \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ d_1 & d_2 & \cdots & d_n \\ f_1 & f_2 & \cdots & f_n \end{bmatrix} \quad (6.3)$$

$$BP = \begin{bmatrix} \frac{S_1}{m} & \frac{S_2}{m} & \cdots & \frac{S_n}{m} \end{bmatrix} \quad (6.4)$$

Where m is the path length and:

$$S_1 = (a_1 + d_1 + f_1)$$

$$S_2 = (a_2 + d_2 + f_2)$$

$$S_n = (a_n + d_n + f_n)$$

S_i in equation 6.4 represents a soft skill's value in a player's BP which fluctuates after each choice made by the player.

To conclude, a human approach to generating a behavioural profile through a GSJT contains three stages, as depicted in Figure 6.3. First, the psychometric team initialised the scoring matrices of the SJT. Then, when the narrative modules are implemented into the GSJT, the player will generate their respective paths in-game. The overall is input into a scoring system, as explained previously, to generate the players' behavioural profile.

6.2.2 A Hybrid scoring method approach

In harnessing the power of the GSJT stealth assessment's capability to amass large-scale, intricate data on players' behaviours, we present **a hybrid proposal**: fusing the **traditional human-centric methodology**, predicated on the scoring matrices derived in the methodology's secondary stage, with a **computational psychometric model**. This model draws upon a resource pool of testers who have **pre-established behavioural profiles**.

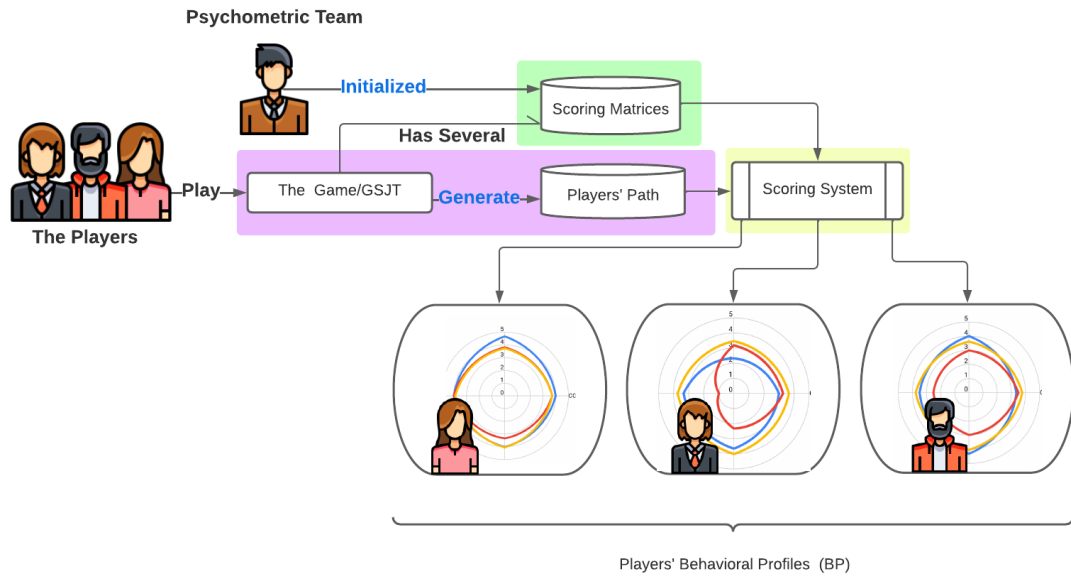


FIGURE 6.3: Psychometric Team Approach to Generate a Behavioural Profile

This amalgamation of the conventionally human-focused approach, where psychometric experts identify the response vectors (RV) with a computational psychometrics scoring matrix system, presents an alternative solution. It appears to incorporate the beneficial elements from both ends of the spectrum, providing **psychometrics experts with an enriched understanding and control over the GSJT's capacity to analyse behavioural profiles.**

The computational psychometrics scoring system capitalises on the principle of **proportion consensus and weighted scoring approaches.** This technique exploits **the players' game paths and their associated pre-identified behavioural profiles to efficiently generate weighted scoring matrices.** Consequently, this system proffers a data-informed methodology for assessing expertise, as opposed to relying purely on subjective human observations or evaluations.

By striving to implement this approach, our objective is to furnish a tool that expedites the creation of scoring matrices. This production, coupled with the insights gleaned from computational data, equips psychometrics experts with a comprehensive perspective. Hence, the proposed model presents an advantageous blend of machine learning approaches and human intuition, potentially enhancing behavioural assessment through a GSJT.

As illustrated in Figure 6.4, the proposed alternative methodology encompasses **four distinct stages.** The initial stage involves **the identification of each**

participant's behavioural profile. Depending on the context, the generation of these behavioural profiles can be achieved either through direct self-assessment utilising the Situational Judgment Test (SJT) devised by the psychometric team, which serves as a behavioural profile SJT, or through a self-peer assessment method [220, 229].

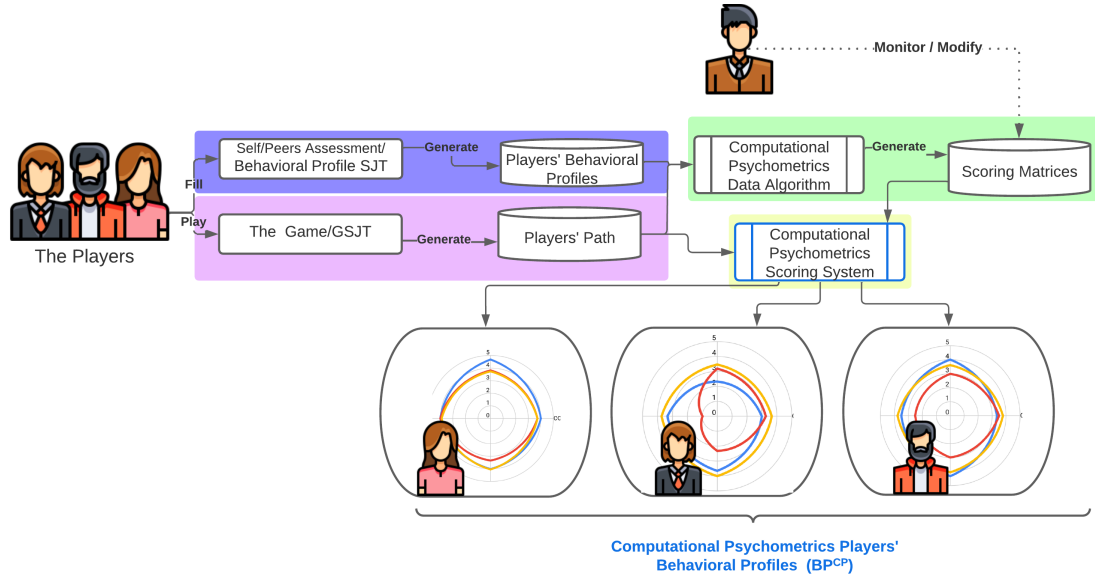


FIGURE 6.4: Computational Psychometrics Approach to Generate a Behavioural Profile

The subsequent three stages share common objectives with the human approach. However, the generation of scoring matrices in this alternative strategy can be done without the involvement of psychometric experts. Instead, the scoring matrices are produced via a computational psychometric data algorithm that considers the participants' behavioural profiles and respective players' paths, as delineated in Figure 6.5. However, our objective is to promote a hybrid approach that capitalises on the advantages of both computational and human methods. Thus, psychometric experts are empowered **to monitor or modify the scoring matrices**.

The computational psychometric data algorithm generates a **two-dimensional array for each node or situation's alternative**, wherein each row signifies the behavioural profile of the player who selects the corresponding situation's alternative. For instance, the two-dimensional array associated with node *A* comprises two rows, while the array for node *F* consists of three rows. Subsequently, an array is constructed using the means of each column in the two-dimensional array, representing the computational psychometric response vector. This process is

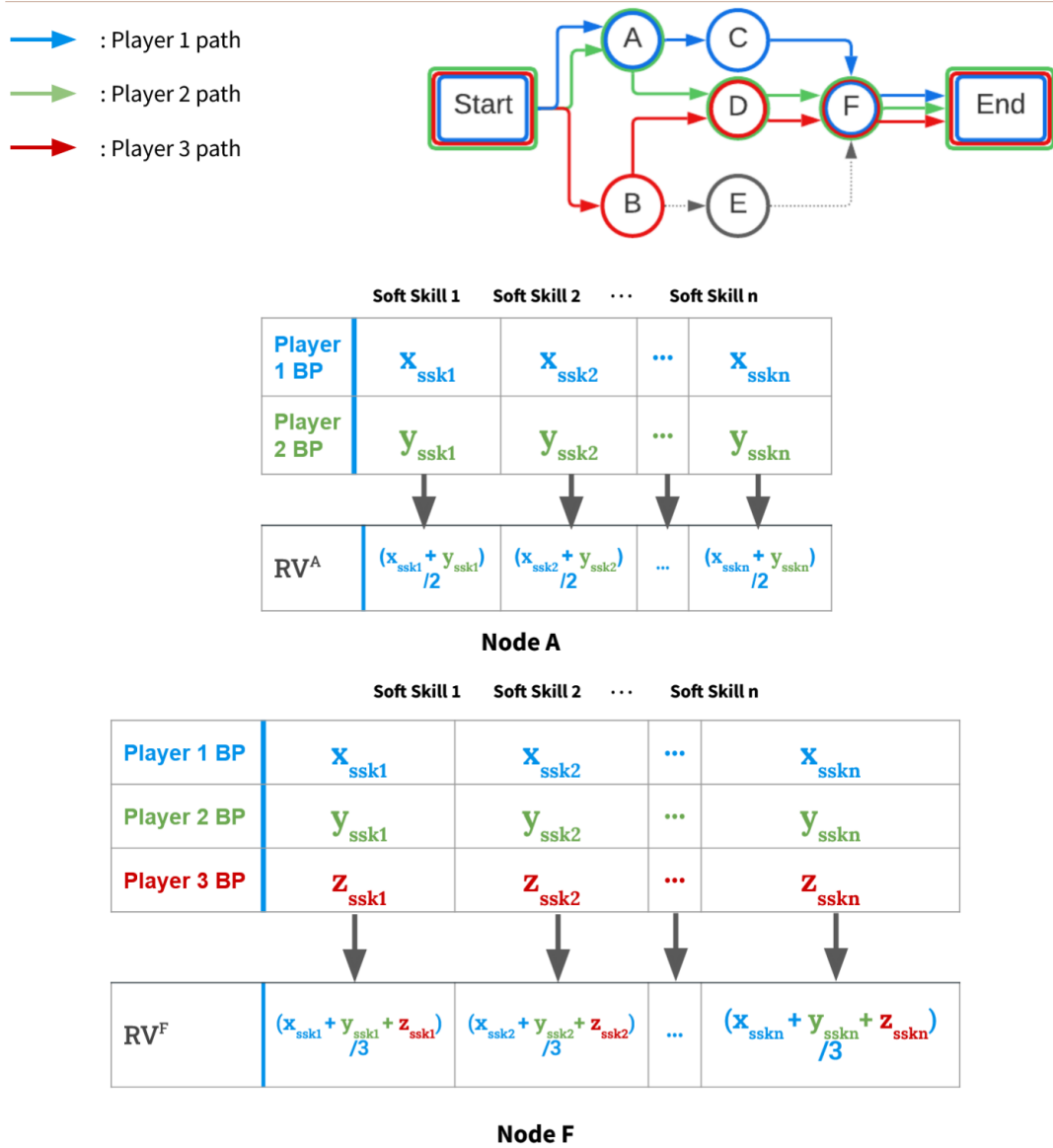


FIGURE 6.5: Computational Psychometrics Algorithm Process

replicated for all narrative module nodes. Consequently, as more players choose a response alternative, the alternative response vector will be increasingly tuned.

To conclude, the primary objective of using this **hybrid approach** is twofold. First, it is not designed to replace human expertise but **to facilitate the rapid creation of scoring matrices**. Second, it aims to offer psychometric experts **new insights through the use of computational data**. These objectives are achieved based on an implemented narrative module. This module leverages the **GSJT stealth-assessment capability** to collect large-scale and complex data on players' behaviours.

6.3 Experimentation 3: Computational Psychometric Scoring System Implementation, Experimental Design and Objectives

In this section, we will present **an experimental study to investigate the effectiveness of using computational psychometrics (CP) in the scoring process** for a gamified situational judgment test designed to assess leadership soft skills. We will compare the results of two scoring methods: **one based on psychometric expert judgment to create the narrative modules scoring matrices response vectors (RV^{expert}) and the other on computational psychometrics (RV^{CP}), using data collected from players' self/peers assessments.** We aim to evaluate whether CP can help psychometric experts in their endeavours to generate response matrices for the situations.

In the context of recent research promoting the use of machine learning in behavioural science [230, 231], we propose to confirm the following hypothesis:

- **The application of computational psychometrics may assist psychometric experts in developing a relevant in-game soft skill measurement system.**

6.3.1 Experimental Gamified Situational Judgment Test

The previously developed *Flight GSJT* (see section 4.3) was adapted for the current experiment to exclusively assess four vital leadership soft skills within an academic context: (1) Facilitating a shared understanding of collective goals (2) Fostering cross-functional collaboration and cooperation, (3) Building trust and rapport, and (4) Encouraging innovation and organisational agility. Players are exposed to **six designed situations** selected from a collection of nine possible scenarios within the game. Four of these six scenarios are prearranged to be experienced by all participants to ensure a comprehensive evaluation of each player's leadership abilities. Individual players' remaining two scenarios depend on their specific choices and actions in preceding situations, thus offering a personalised and dynamic experience.

Each scenario presents players with **two or three unique options**, leading to **twenty-five potential alternatives** throughout the game. This framework enables **2430 distinctive pathways**, providing a rich environment for leadership soft skills. In alignment with our assessment objectives, we have removed any form of feedback for this experiment to mitigate potential biases that may influence the evaluation process. Finally, one psychometric expert generated the various narrative module scoring matrices.

6.3.2 Sample and Procedures

Participants for the GSJT were recruited from a diverse range of nine companies spanning Canada, Italy, Germany, Turkey, and France. These participants represented fifteen separate working teams, totalling **seventy-six players**. Team sizes varied, consisting of **three to nine members** with a combination of supervisors, subordinates, and peers of equal hierarchical status.

After completing the GSJT, players were asked to complete **a survey to assess themselves and their peers' performance** in the four targeted soft skills. This assessment utilised the rating scale outlined in Table 6.2, which experts employed to score situation alternatives. Players could abstain from evaluating an individual or a specific soft skill if they felt inadequately informed about the person or the skill in question. **A minimum of two peer assessments** was mandated to guarantee fairness and equity in generating a player's behavioural profile from the peers (BP^{peers}). As a result, **forty** BP^{peers} were generated from the initial fifty-one players.

TABLE 6.2: Scoring Grid

Score	Description
5	Extensive experience in the soft skill area
4	Good experience in the soft skill area
3	Some experience in the soft skill area
2	Little experience in the soft skill area
1	No experience in the soft skill area
0	The alternative does not evaluate the soft skill

Upon analysing the gameplay data, we observed **thirty-six unique paths taken by the fifty-one players** who completed the game. In total, **20 out of the 25**

potential alternatives were encountered throughout the various playthroughs. Consequently, twenty RV^{CP} were generated.

Ultimately, **three distinct types of BP**—each representing a player’s behavioural profile were associated with each participant: (1) **An expert behavioural profile (BP^{expert}) based on scoring matrices established by the expert**, (2) **a computational psychometrics behavioural profile (BP^{CP}) derived from the twenty RV^{CP}** , and (3) **the reference self/peers behavioural profile (BP^{peers})**.

6.3.3 Results and Discussion

The proposed hypothesis suggests that incorporating a computational psychometrics (CP) scoring system **may assist psychometric experts in developing a relevant in-game soft skill measurement system**. This enhancement is achieved by generating CP scoring matrices, as detailed in the corresponding method section 6.2. To investigate this, we conducted a **similarity analysis involving the players’ BP^{expert} , their BP^{CP} , and their BP^{peers}** .

In order to assess the reliability of the CP scoring system, we carried out an **Euclidean distance (L2) analysis** between the players’ BP^{CP} and BP^{peers} . This analysis offers insights into the level of agreement between the BP^{CP} and BP^{peers} . A smaller distance would indicate higher similarity and, therefore, increased reliability.

Similarly, we also performed an **L2 analysis** comparing the players’ BP^{expert} and BP^{peers} to evaluate the reliability of the expert scoring system. By comparing the results of these two analyses, we can determine how much the CP scoring system enhances the GSJT’s scoring system reliability compared to the expert scoring system. This comprehensive examination enables us to understand better the potential benefits and drawbacks of incorporating computational psychometrics into the GSJT scoring system.

We found **superior similarity results for the L2 norm between players’ BP^{CP} and BP^{peers} (Average = 0.92, Median = 0.90, Min. = 0.32, Max. = 1.57)** compared to the players’ BP^{expert} and BP^{peers} (Average = 1.76, Median = 1.68, Min. = 0.63, Max. = 3.34). Figure 6.6 graphically illustrates the distinct behavioural profiles and demonstrates that an L2 norm of 0.90 signifies that

the two scoring systems generate nearly identical score values for each soft skill. This supports our hypothesis that **a CP scoring system may be reliable for evaluating behavioural profiles**. Conversely, an L2 norm of 1.68 indicates a discrepancy in the values of three soft skills, suggesting that the expert scoring system is less precise in its assessment.

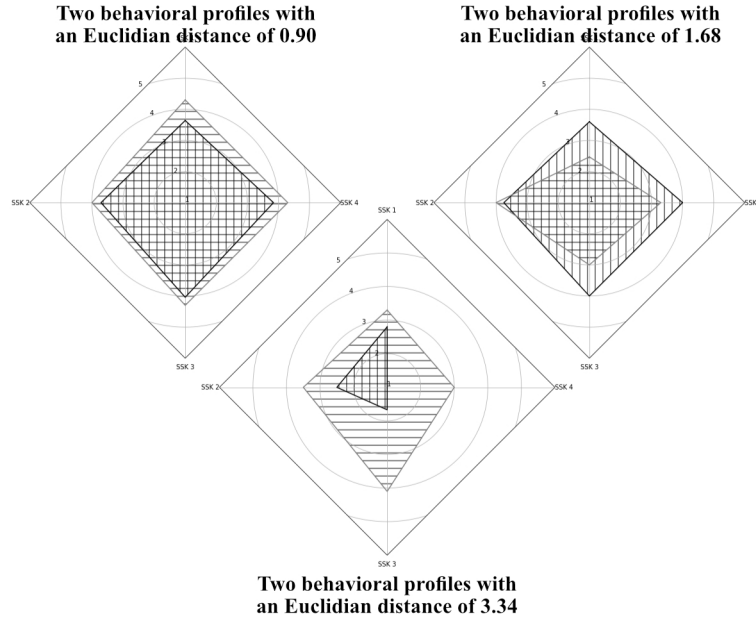


FIGURE 6.6: Euclidian analysis between three pairs of behavioural profiles

We also conducted **a cosine similarity analysis**, which revealed a high similarity with an average of **0.99 between BP^{CP} and BP^{peers}** . These findings further support our hypothesis. Figure 6.7 presents the distribution, spread, and skewness of the L2 norm analysis using a box-plot method to delve deeper into the data and interpret the results. The four guidelines' level numbers represent the soft skills the two scoring systems disagree with in the BPs. The two scoring systems concur if the L2 norm between the results is under the level 1 guideline. Conversely, above guidelines 1, 2, 3, and 4, the scoring systems disagree on at least one, two, three, and four soft skills, respectively, as indicated in table 6.2. For level 4, this may also imply disagreement on a single soft skill with **a significant scoring difference of (≥ 2.0)**.

First and foremost, Figure 6.7 demonstrates that the L2 analysis of the expert scoring system is much more dispersed than that of the CP scoring system, with a distribution **centred around level 3**. This suggests that the expert scoring system's assessment is less consistent with the self/peer assessment than the CP scoring system's assessment is with the self/peer assessment. Furthermore, we

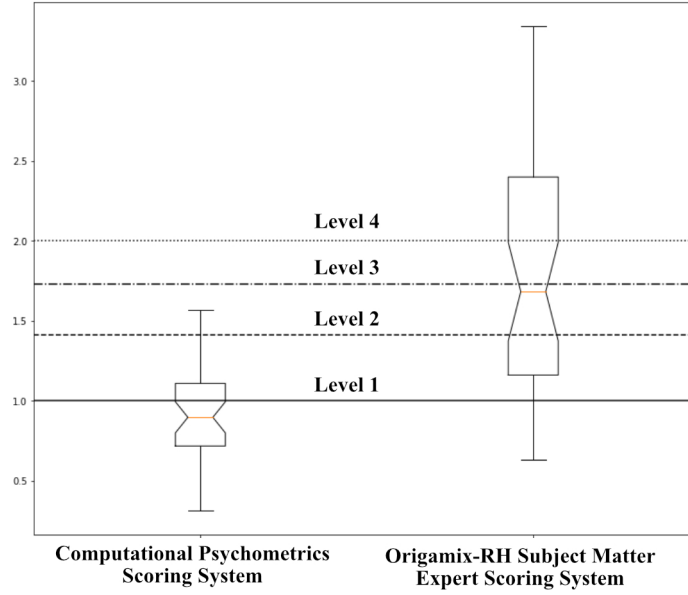


FIGURE 6.7: Euclidian Analysis Box-plot Visualization for the Both Scoring System

observed that **20% of BP^{expert}** are in complete agreement with BP^{peers} (under the guideline level 1), whereas the CP scoring system revealed that **70% of the BP^{CP}** are in complete agreement with the BP^{peers} . As a result, the findings indicate that CP is more reliable than the expert scoring system, with a high level of reliability in assessing soft skills.

Therefore, **a CP scoring system could be used as an alternative scoring system for GSJT**. A computational psychometrics scoring system might be more straightforward and efficient, as obtaining RV^{expert} scores from an expert can be time-consuming and cumbersome. By combining game sessions and data collection through the digital GSJT, it is possible to generate RV more efficiently, accurately, and reliably than through a unique expert scoring approach. Computational psychometrics may also reduce cognitive errors and biases. For instance, the expert realised after the experiment that they had confused two characters' names in the game story, which affected the situation's scoring matrix. The expert also admitted that working closely with the game client influenced their scoring, implying that the expertise needed to be more neutral and more aligned with the client's perspectives on which situation answer was good, deviating from the initially desired honest assessment approach.

However, this innovative approach is not without its challenges. The inherent data-driven nature of computational psychometrics brings with it potential biases. For

example, **the selection of the initial player pool**, critical to initialising scoring matrices, may introduce inherent bias based on the pre-established reliability of players' behavioural profiles. Hence, the selection and design of the initial player pool require careful consideration to generate reliable and unbiased scoring matrices. As outlined in the chapter 1, despite their inherent biases, experts should remain an integral part of the process due to their profound understanding of the subject matter and context. Moreover, the limitations of the computational psychometric system demand **continuous expert monitoring to mitigate any shortcomings**. Hence, our hybrid approach aims to harness the advantages of both approaches, **reducing inherent biases through their complementary interaction**. Yet, human resources experts may require additional support to fully rely on computational data methods, leading us to explore whether a CP scoring system can serve as an effective tool in aiding expert scoring.

To this end, we asked the expert to repeat the scoring process while considering the CP scoring for each situation's alternatives. We observed that the expert altered **thirteen of their RV expert scores out of the twenty alternatives attributed to an RV^{CP} by the CP scoring system**. Nonetheless, the expert's modifications only affected some alternative RV^{expert} , which did not initially consider evaluating one or more soft skills with the modified RV^{expert} . We observed that the expert changed their mind about their initial alternative scoring of specific soft skills, ranging from 0, as described in table 6.2, to a higher value. However, the expert's feedback on their second scoring round indicated they only **lightly considered the RV^{CP} in the iteration**. The expert also felt that the CP scoring system needed to fully account for the human and social complexity of the game's situations. In other words, the modifications were not prompted by the CP scoring system but rather by the effect of iterating the alternatives' scoring.

Conversely, a second external expert's scores, obtained after playing the game, aligned closer to the CP scores. This expert also utilised **the CP scores to refine scoring matrices and fully trusted the computational scoring system for four scoring alternatives**.

These observations demonstrate **a variance in experts' utilisation of computational data methods**. Some experts might use the CP scoring model as **an auxiliary tool** to partially or fully refine their scoring methods. In contrast, others might **fully trust the computational psychometrics method** and directly incorporate CP scores into their evaluations.

In conclusion, our findings suggest that the CP scoring model could serve as a valuable tool for experts in soft skills assessment. However, further research is necessary on the one hand to enhance and minimise the potential biases of the computational scoring system and make it valid with other samples and studies; on the other hand, to understand experts' readiness to adopt computational data methods in their scoring processes. One approach could be the application of the Technology Acceptance Model. This model can help explore the factors influencing experts' acceptance and use of computational psychometrics in their scoring processes, such as perceived usefulness, perceived ease of use, and subjective norms. Conducting more thorough research in this area can provide insights into potential barriers and facilitators influencing experts' willingness to integrate computational data methods. This, in turn, could contribute to developing more effective and reliable scoring systems for evaluating soft skills in GSJTs.

6.4 Conclusion

In conclusion, our research indicates that **the computational psychometrics scoring approach could be a potent instrument for experts evaluating soft skills**. The computational psychometrics scoring approach and the presented experimentation have been the subject of a peer-reviewed scientific paper [232]. Nonetheless, further exploration is necessary for two main areas. Firstly, to amplify the effectiveness of the computational scoring system by minimising its potential biases and validating its applicability with diverse samples and studies. Secondly, to delve deeper into understanding the readiness of experts to incorporate computational data methods into their scoring procedures.

A more exhaustive investigation in this realm could offer valuable insights into the potential hurdles and enablers that influence experts' willingness to incorporate computational data methodologies. Ultimately, these research efforts could contribute to the evolution of scoring systems for GSJTs. By harnessing the power of computational data methods, we aim to enhance the effectiveness and reliability of soft skills assessment, thereby potentially creating a more accurate and nuanced understanding of individuals' soft skills. Thus, **the potential of computational psychometrics, combined with experts' proficiency, presents a promising future for the advancement of assessment processes**.

Chapter Overview

In alignment with **research objective B**, we delved into the development of a **computational psychometric scoring system** in this chapter, designed to aid the psychometrics teams of GSJT.

In the first part, we explore **the state of the art of the scoring mechanisms** involved in situational judgment tests (SJT). Traditional scoring methods are outlined, each with distinct modalities, benefits, and limitations. The goal is to establish context by providing a foundational understanding of the current landscape in SJT scoring.

Transitioning into the second part, we introduce **our proposed solution - a computational psychometric scoring approach that incorporates traditional GSJT human scoring methods**. This approach leverages scoring matrices and machine learning techniques for behavioural profiling. We suggest **an hybrid scoring method to build on the strengths of human-centred traditional systems while counteracting their weaknesses**. This method combines the precision of human-centred computational scoring with the familiarity and comfort of traditional methods and the expertise of a psychometrics expert. The aim is to offer a human-centred, unique, and nuanced evaluation of an individual's responses.

The latter part of the chapter is dedicated to a human-centred practical application of our proposed system and the **hybrid approach**. We conclude the chapter with the results of our experiment and a discussion of their implications. Through discussing the performance of our computational scoring system, we aim to demonstrate its feasibility and potential benefits over the design and production of GSJTs.

Chapter 7

Machine and deep learning models to help the creation of realistic non-playable characters

7.1 Why it's important to have realistic characters for soft skill recognition

As explained earlier, interactive narrative games are a primary form of **story-telling that relies heavily on interpersonal situations with various characters**. As a result, one of the primary reasons that realistic NPCs with real personalities are crucial in a GSJT is that they contribute to **the overall immersion and engagement of the participant**. Immersive experiences are more likely to **elicit genuine responses** and encourage participants to **draw upon their accurate soft skills** in order to navigate complex social situations. When NPCs display authentic emotions, reactions, and behaviour, they could create an environment where the participant can fully immerse themselves, ultimately providing a more accurate reflection of their soft skills. Moreover, heightened immersion helps maintain participant **motivation and interest**, reducing the risk of disengagement or inaccurate responses due to boredom or lack of investment in the test.

In the field of gamification, it is a commonly held view among experts that NPCs, more specifically in ING, should ideally capture the nuances of human behaviour

to effectively stimulate a participant's decision-making skills [233, 234]. It is proposed that realistic NPCs, equipped with well-defined personalities, may encourage participants to consider diverse factors in decision-making processes, such as **the character's emotions, motivations, and interpersonal dynamics** [210, 235]. It is argued that the complexity introduced through these realistic NPCs could mirror the intricate nature of decision-making in real-world scenarios, thereby facilitating a more authentic evaluation of soft skills.

Moreover, our expert collaborators hypothesise that these life-like NPCs can potentially enhance **the transference of soft skills learned and honed** during the GSJT. By engaging with NPCs that strive to **simulate natural human behaviour**, participants and, consequently, recruiters may be better positioned to identify and select individuals with key soft skills that are likely to be beneficial in their future everyday lives. It is also suggested that these realistic NPCs might serve as **a bridge between the simulated GSJT environment and the actual world**, potentially bolstering the development and recognition of soft skills across various contexts. It is, however, important to note that empirical research to support these hypotheses conclusively is currently limited, pointing towards an interesting direction for future studies.

In conclusion, integrating realistic, non-playable characters with true-to-life personalities into gamified situational judgment tests may be key for authentic and reliable soft skill evaluation. These characters could foster immersion, encourage complex decision-making, and potentially cultivate empathy and emotional intelligence, in addition to enhancing the transferability of skills to real-life scenarios [233]. As the need for efficient soft skill evaluation tools continues to rise, including realistic NPCs in the design and deployment of gamified situational judgment tests remains an area of significant interest, both for its current perceived benefits and future empirical investigation.

7.2 Proposing dialogues personality classifiers

Following our methodologies, **the writers** hold the second most crucial role in **conveying the coherence of the assessment**, as they are responsible for transforming the situational judgment test (SJT) situations into an **interactive narrative script**. A fundamental process at this stage involves the creation of realistic

characters and dialogues between these characters, as detailed in Section 7.1. The psychometric team validates this process, ensuring the intended SJT situations are accurately transposed into an interactive narrative script.

In alignment with our objectives to support the GSJT writing team in their efforts, we propose **the development of personality classifiers based on natural language processing (NLP) that will assist the writers**. These classifiers will help to facilitate the creation of realistic characters and dialogues, further enhancing the gamified situational judgment test as detailed in section 7.1.

Building on the approaches of Rahman et al. [236] and Majumder et al. [237], personality traits individually yield superior results compared to studying the combination of traits as a single type. Several methodologies exist for generating binary classifiers for personality recognition, which primarily rely on **machine learning and deep learning algorithms** that are heavily data-driven.

In this context, the objective is **to compare various speech-based binary classifiers for character personality recognition**, focusing on one or multiple typology models, such as the four *MBTI* dichotomies and the *BigFive* traits (see Section 1.1.4). This comparison enables the recommendation of classifiers best suited to the specific context of writers and developers. However, before comparing these classifiers, we propose a method to facilitate the creation of such classifiers in the realm of non-playable character personality recognition.

7.3 Creation of a binary classifier for Non-Playable Characters Dialogues

Whether the classifier is based on machine learning or deep learning techniques, we suggest using **dialogues/speeches and characters' psychometric evaluations** from existing INGs as input data for training the models. The proposal procedure for creating classifiers in natural language processing for NPC personality recognition is described in three stages in Figure 7.1: **Document collection and labelization, document preprocessing, and training and validation of the generated models**.

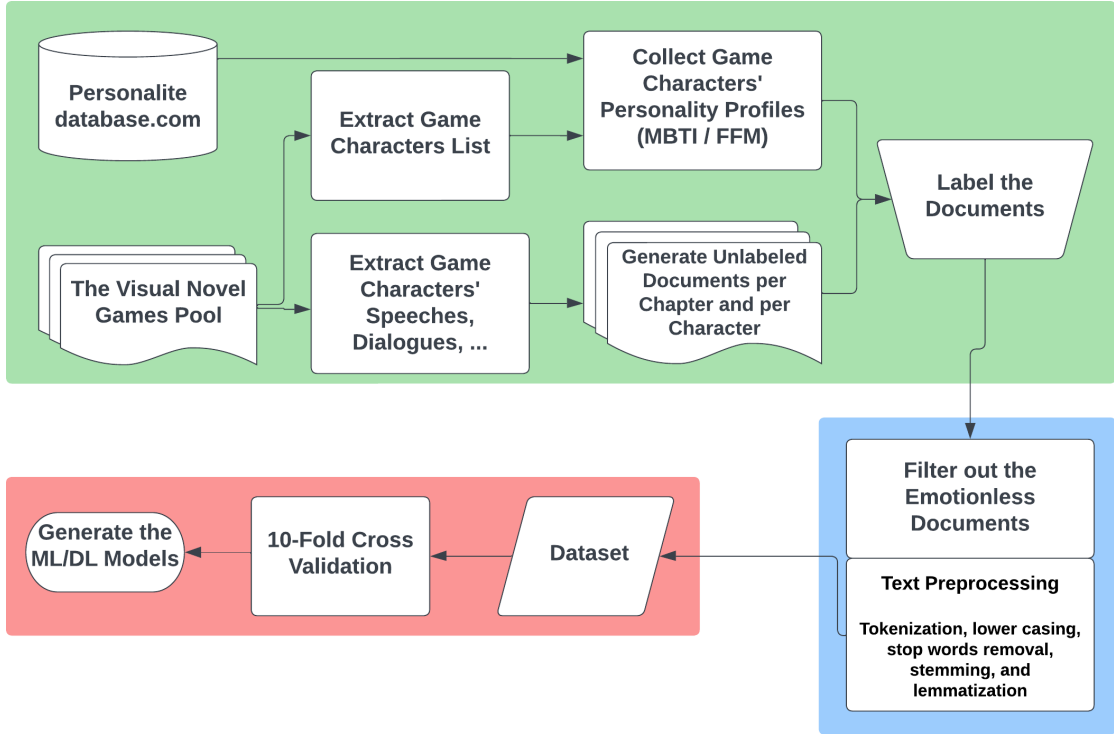


FIGURE 7.1: Personality Classifier Generation Process

7.3.1 Stage 1: Document collection and labelization

Firstly, the GSJT designers have to **collect a maximum of games** where the characters and their respective dialogues are known to be used as input for the training of the supervised machine and deep learning models. In supervised machine learning, using labelled data as input is essential for effectively training the model. Consequently, each character must be tagged with **multiple labels that correspond to the selected model's personality traits**. We believe players are best suited to tag these characters since they directly engage with the NPCs throughout the game. To facilitate this process and follow our research objectives, we used the "*personalitydatabase.com*" platform, which allows players to vote on the personality traits of various fictional characters from movies and video games. This community-driven website primarily focuses on **the Myers-Briggs Type Indicator (MBTI) and the Big Five personality typologies**.

While characters' feelings and emotions might change during a game based on the storyline, a character's overall personality, as described by MBTI or the Big Five, remains consistent throughout the entire game. Players' perceptions of a character's personality typically do not change significantly from beginning to end unless the story involves a character with a split personality disorder, such as in the

Danganronpa series games. In cases like these, we treat the different personalities as separate characters.

Moreover, we believe that each game chapter¹ serves as a unique context in which characters' personalities are exhibited in ways corresponding to the specific chapter situation. Compiling all dialogues and speeches of a character into a single document could potentially confuse the training process, as the character may express a wide range of emotions throughout the game. Therefore, we also propose to use deep learning techniques to capture the emotional behaviour of a character in a particular situation while considering their personality as a constant factor. In other words, **a character's emotional expression may vary depending on the context and situation, but their personality remains constant throughout.**

To address this concern, we propose dividing speeches and dialogues into multiple documents, each representing all the sentences spoken by a specific character within a particular game chapter. This approach allows for a more comprehensive and nuanced understanding of each character's personality traits and emotional expressions, ultimately leading to better performance in training the model. Finally, we are able to label each document with multiple labels based on the respective characters' personalities.

7.3.2 Stage 2: Document Preprocessing

Binary classifiers aim to categorise textual data into one of two predefined classes (e.g., introverted or extroverted). For such classifiers to be effective, it is essential to filter out emotionless documents and perform thorough text processing.

Emotionless documents contain little or no emotional content, making it difficult for NLP models to discern personality traits accurately. Including such documents in the dataset used for training and evaluation may lead to a biased or less accurate classifier, as the model may fail to learn the relevant patterns and relationships between textual features and personality traits.

¹A *chapter* in the context of interactive narrative games typically refers to a segment or part of the game's overall storyline. Chapters are commonly used to divide the game's narrative into manageable portions, each with its own subplot or set of objectives contributing to the overall plot.

Additionally, text processing is a crucial step in developing any NLP model. In the context of binary classifiers for personality detection, the following text-processing techniques are particularly relevant:

- **Tokenization:** Tokenization involves breaking a document into individual words or tokens. This process allows the model to analyse the frequency and distribution of words, which can be informative for identifying personality traits.
- **Stopword Removal:** Stopwords are common words (e.g., "the," "and," "is") that carry little meaning and can introduce noise into the model. Removing stopwords helps to focus the model's attention on more relevant and informative words.
- **Lowercasing:** Converting all text to lowercase ensures consistency and reduces the dimensionality of the data, as words in different cases are treated as separate tokens.
- **Lemmatization/Stemming:** Lemmatization and stemming are techniques for reducing words to their base or root form. This process helps to group similar words together, reducing noise and enabling the model to identify patterns more effectively.
- **Feature Extraction:** Transforming raw text into numerical representations, such as word embeddings or term frequency-inverse document frequency vectors, allows the model to analyse and process the data more efficiently. These representations capture semantic and syntactic information, which can be beneficial for personality detection tasks.
- **Feature Selection:** Reducing the number of features used in the model can help prevent overfitting and improve generalisation. Techniques such as mutual information, chi-squared, or recursive feature elimination can be used to select the most relevant features for the task.

In conclusion, developing a binary classifier for personality detection using NLP requires careful consideration of filtering emotionless documents and thorough text processing. By eliminating documents lacking emotional content, we ensure that

the classifier is trained on a more representative and informative dataset, improving performance and accuracy. Furthermore, applying text processing techniques, such as tokenisation, stopwords removal, lowercasing, lemmatisation/stemming, feature extraction, and feature selection, enables the model to effectively capture the underlying patterns and relationships between textual features and personality traits.

7.3.3 Stage 3: Train and Generate the Models

Finally, the last stage focuses on training and validating the binary classifiers. As mentioned earlier, several approaches exist based on machine learning algorithms and deep learning algorithms. Our objective in this chapter is not to enhance the validity and reliability of such an algorithm but to use it to generate a useful model to help the writer team in their endeavours. Therefore, we suggest training several models for each personality trait with the most used machine algorithm from the literature review for natural language processing for personality classification: **support vector machines, random forest, logistic regression, naive Bayes and decision tree.**

In addition, following Rahman et al.'s approach [236], **a convolutional neural network (CNN)** proves its efficiency in text-based personality recognition. Furthermore, CNN enables the training of a multi-class classifier. Therefore, following the same three-step process, the authors may generate a multi-class classifier to help them evaluate the personality of their characters.

Lastly, we suggest using **the 10-fold cross-validation** used in machine learning, including personality detection tasks, to assess the performance of a classifier on unseen data. Cross-validation aims to obtain a more accurate and reliable estimate of the classifier's generalisation ability, which helps prevent overfitting and underfitting. Hence, in our context, employing such a validation model may assist writers in selecting or generating the most suitable model for their particular objective. This goal involves converting situational judgment test scenarios into interactive narrative situations incorporating engaging and realistic dialogue between characters.

7.4 Experimentation 4: Development and implementation of NPC’s binary personality classifiers, Experimental Design and Objectives

In this section, we present our approach to experiment with text-based personality classifiers for categorising non-playable characters (NPCs) in visual novels, aimed at assisting game designers and writers in crafting more engaging and interactive gamified selection tools. Building on the ideas discussed in section 7.3, we employ natural language processing techniques and evaluate various machine learning and deep learning methodologies to classify NPCs based on their dialogue.

Our work in this section serves two main purposes. First, we argue that machine learning and deep learning techniques effectively identify distinct NPC personalities. Second, we aim to compare and identify the most suitable algorithmic approaches for generating personality trait classifiers. The primary focus of this section is to gather preliminary evidence in support of the hypothesis that these techniques can successfully categorise NPCs according to their personalities.

Hence, this section focuses on finding preliminary evidence to confirm the following hypothesis:

- **By employing machine learning and deep learning techniques, it is possible to classify non-playable character (NPC) personalities in order to enrich and heighten the significance of in-game interpersonal interactions.**

7.4.1 Psychometry, Tools and Dataset

In this subsection, we first describe the two psychometry typologies used in the experiment (MBTI & Big Five). Then, as Furnham et al. [42] revealed in their study, there is a high correlation between the two. As we will see later, we present a pre-study to check if this correlation may be applied and effective in our context. Finally, we present the tools and the dataset used for the whole experiment.

7.4.1.1 MBTI and Big Five Typologies

Building upon the research by Ochs [238] that aims to create believable NPCs, our study examines two prevalent psycho-social typologies for characterising personalities. First, we consider the Myers-Briggs Type Indicator (MBTI) [42], a widely used assessment tool that measures an individual's personality type based on four distinct dichotomous traits:

- *Extraversion/Introversion (E/I)* dichotomy reveals how an individual interacts with his/her environment and others.
- *iNtuition/Sensing (N/S)* dichotomy reflects how an individual perceives the collected information from his/her surroundings.
- *Feeling/Thinking (F/T)* dichotomy depicts how an individual decides to act.
- *Judging/Perception (J/P)* dichotomy determines the attitude that influences an individual's way of living or working.

Combining the four dichotomous traits, we arrive at 16 distinct personality types. Each type is denoted by a four-letter code encapsulating an individual's unique personality characteristics. For instance, an individual possessing Extraversion, Intuition, Feeling, and Perception traits is classified as an ENFP type. ENFP individuals, often referred to as free spirits, are known for their cheerful and communicative nature. They are deeply connected to their values and continually seek existential harmony.

In contrast, we also explore the Big Five personality factors [42], which consist of five defining traits that remain relatively stable throughout an individual's life. These traits include:

- *Extraversion (EXT)* is related to positive emotions and the tendency to seek stimulation and company from others.
- *Openness to experience (OPN)* is related to appreciating art, emotions, adventures, unusual ideas or new ideas, curiosity, and imagination.
- *Agreeableness (AGR)* is related to being compassionate and cooperative rather than suspicious and antagonistic towards others.

- *Conscientiousness (CON)* is related to self-discipline, respect for obligations, organisation rather than spontaneity, and goal-oriented.
- *Neuroticism (NEU)* is related to a tendency to easily experience unpleasant emotions such as anger, worry or depression, vulnerability, and stress.

Compared to the MBTI's dichotomies that reflect binary oppositions in meanings, the Big Five factors follow a continuum logic for each trait. In other terms, an individual is more or less presenting aspects that define a Big Five trait. However, in this study, we consider the Big Five traits as boolean variables to enable the generation of binary classifiers following the related works' approaches.

7.4.1.2 Pre-study on the Correlation between the Big Five and MBTI typologies

Furnham and Adrian [42] conclude in 1996 that there is a correlation between the MBTI indicators and the Big Five personality traits [239]. Table 7.1 depicts the Correlation between the two typologies.

TABLE 7.1: Correlation between the Big Five traits and the MBTI traits

MBTI	Big Five
E/I	EXT (correlates with E)
N/S	OPN (correlates with N)
F/T	AGR (correlates with F)
J/P	CON (correlates with J)
Inexistent in MBTI	NEU (No Correlation)

However, we need to know if this Correlation may be applied and effective in our context. Hence, we first measure the inter-rater agreement reliability of each character's personality between the two typologies using Cohen's kappa coefficient approach. Two methods were suggested:

- One method measures the inter-rater agreement coefficient by personality type. As the MBTI typology does not have the neuroticism trait from the Big Five, we measure the coefficient using only the correlated traits in Table 7.1.

In other terms, we have a perfect agreement if a character was labelled as an ENFJ type and presented the extraversion trait, the openness to experience trait, the agreeableness trait, and the conscientiousness trait of the Big Five typology.

- The second method measures the correlated traits independently of the personality types. Therefore, we measure four Cohen's kappa for each correlated trait combination to validate inter-rater agreement reliability.

The inter-rater agreement pre-study reveals a moderate agreement of **0.589** for the first method, revealing the overlap between the personality types of the two typologies. Likewise, the second method reveals a correlation between each proposed correlated trait in Digman and John's study [239]. Nonetheless, as shown in Table 7.2, the Correlation varies from a moderate agreement for the openness to experience and conscientiousness traits with their MBTI pair-wise traits through a substantial agreement for the agreeableness trait with its MBTI pair-wise traits to a near-perfect agreement for the extraversion trait with its MBTI pair-wise traits.

TABLE 7.2: Interpretation Cohen's Kappa value

Pair-wise traits	Kohen's Kappa	Interpretation
EXT - E/I	0.816	Near-perfect agreement
OPN - N/S	0.467	Moderate agreement
AGR - F/T	0.664	Substantial agreement
CON - J/P	0.523	Moderate agreement

These results are consistent with Digman and John's results [239], which suggest an overlap between the two psychometric typologies. However, the variation between the pair-wise correlated traits suggests that there is still a difference between them. For this reason, we will study both typologies independently for the rest of the study.

7.4.1.3 Experiment Tools

Our study employs Python as the primary language for data collection, processing, and analysis. We utilise the Numpy Library and the Natural Language Toolkit

(NLTK) for matrix computation and text preprocessing to ensure efficient data manipulation and processing. Furthermore, we implement standard machine learning algorithms using the sci-kit-learn library while leveraging the spaCy library for deep learning applications. This comprehensive approach allows us to explore and compare a range of techniques in our quest to develop effective personality classifiers for non-playable characters.

7.4.1.4 Game Pool & Dataset

We extracted character dialogues from nine renowned visual novels to create our dataset. These games were selected based on the breadth of their character roster and the richness of the personalities depicted within. The chosen games include:

- *Danganronpa: Trigger Happy Havoc* is a Japanese visual novel and murder mystery game featuring a diverse array of characters with distinct personalities. This game marks the first instalment in the Danganronpa series.
- *Danganronpa 3: Ultra Despair Girls* is a sequel to the original Danganronpa, combining third-person adventure shooter elements with visual novel gameplay. The game introduces a new set of characters distinct from those in the series' first episode.
- *Steins;Gate* is a captivating visual novel that follows various characters navigating multiple alternative timelines. The characters' personalities remain consistent while the contexts and character variants change across timelines.
- *Umineko no Naku Koro Ni I & II* are visual novels that share a similar narrative structure with the first Danganronpa instalment. The story features eighteen characters, each with their own unique personalities.
- *Phoenix Wright: Ace Attorney* is a series of visual novel adventure games that showcase a wide range of characters and personalities throughout multiple episodes. Players assume the role of an attorney tasked with defending various clients by investigating cases and presenting evidence during trials.
- *Persona 3, 4, and 5* are role-playing games that incorporate extensive dialogue-driven gameplay. These instalments offer a rich assortment of characters and personalities, blending visual novel elements with traditional role-playing adventure gameplay.

Following our proposal in section 7.3, labelled data are required as input in supervised machine learning to train the model. Hence, each character must be tagged with multiple labels corresponding to the four MBTI dichotomies and the Big Five. The players are the most predisposed to tag the characters as they directly interact with the NPCs. Therefore, we used the "*personalitydatabase.com*" [240] platform that enables players to vote for the personality of any fictional characters from movies or video games. We extracted 170 non-playable characters' personalities from the game pool and labelled each character's speech with their corresponding personalities. We regroup the speeches and dialogues into several documents that represent all the sentences used by a specific character in a specific game chapter. This process provides 6200 labelled entry points in the form (document, MBTI/Big Five traits).

7.4.2 Experimental Method

In line with the approaches employed by Rahman et al. [236], Majumder et al. [237] and our proposal, analysing and examining personality traits independently yields better results than studying the combination of traits as a single type. Consequently, we propose comparing various character speech-based binary classifiers for each of the four MBTI dichotomies and the Big Five traits. Additionally, we utilise deep learning algorithms to generate and compare a multi-class classifier.

We use the dialogues/speeches and the characters' psycho-metrics as the data entry points to train the models. The process of training classifiers in natural language processing is globally divided into three steps:

- Dataset Analysis
- Data Preprocessing
- Model Training

The following subsections of this section explain the details of those steps.

7.4.2.1 Dataset Analysis

We conducted a preliminary analysis of the distribution of MBTI personality types and the various personality traits independently to verify the suitability of the

collected dataset for our research. As shown in Figure 7.2, the dataset is not uniformly distributed across the sixteen personalities defined by the MBTI typology. For instance, the least common personality type among the data points is INFJ, with only 35 occurrences, while the most frequent personality type is ISFJ, with as many as 936 occurrences. This uneven distribution should be considered when interpreting the results and evaluating the performance of the classification algorithms.

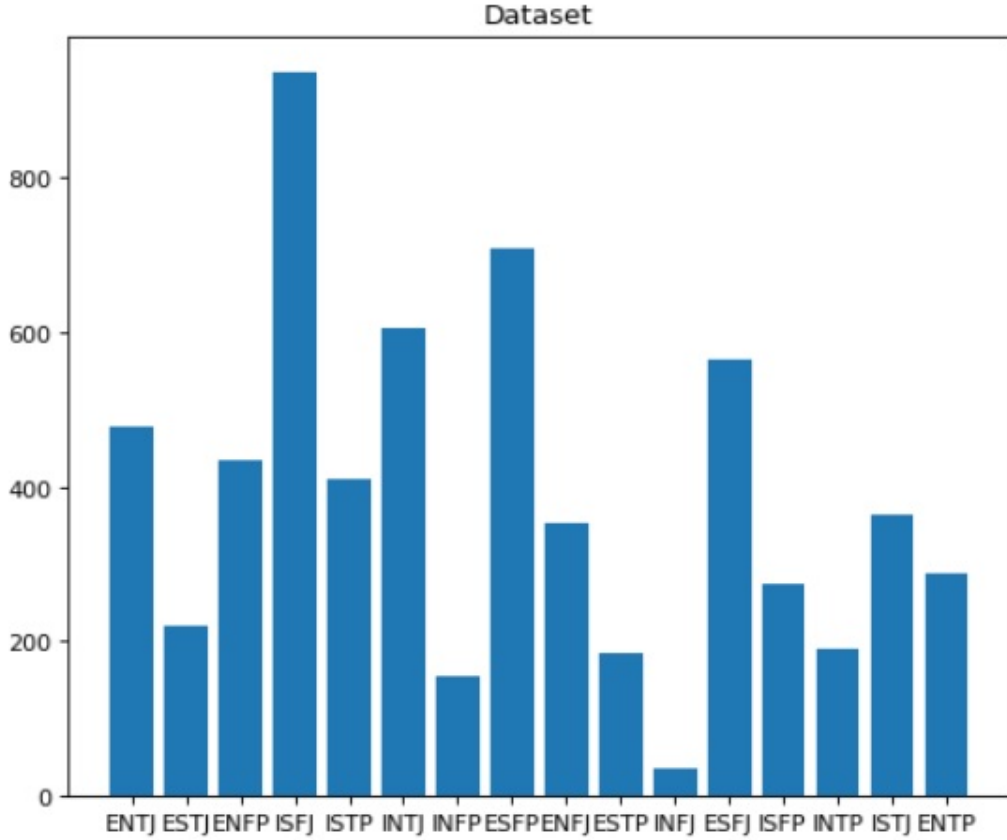


FIGURE 7.2: MBTI Dataset Distribution Analysis

Despite the uneven distribution of MBTI personality types, our study aims to compare binary classifiers for each personality trait, following the methodology of Majumder et al. [237]. As shown in Table 7.3, the data is more evenly distributed among the four MBTI dichotomies when considered independently. The difference between the two possible labels for a specific trait is not significant enough to warrant remediation methods such as class duplication or reduction [241]. Consequently, we proceed with training on the original dataset, which consists of 6,200 data points for the MBTI dataset, and carefully consider the distribution when interpreting results and assessing classifier performance.

TABLE 7.3: MBTI's Traits Distribution

MBTI Personality Trait	Occurences
E / I	3231 / 2969
S / N	3663 / 2537
T / F	2736 / 3464
J / P	3556 / 2644

A similar observation holds for the study of Big Five personality types. Since the Big Five typology includes a fifth personality trait, there are thirty-two possible personality types. However, only 30 of the 32 types are present in our dataset. For this reason, as mentioned in the methodology section, we will not explore the creation of a multi-class classifier for the Big Five typology. Instead, Table 7.4 indicates that we can independently focus our analysis on the Big Five traits. It is important to note that the dataset for the Big Five consists of 5,686 data points, as players labelled fewer characters with the Big Five traits than the MBTI dataset.

TABLE 7.4: Big Five's Traits Distribution

Big Five Trait	Occurrence
EXT / \overline{EXT}	3009 / 2677
OPN / \overline{OPN}	3459 / 2227
AGR / \overline{AGR}	3646 / 2040
CON / \overline{CON}	2872 / 2814
NEU / \overline{NEU}	3148 / 2538

7.4.2.2 Data Preprocessing

In order to train our models effectively, we filter out documents of emotionless characters that do not convey any personality cues, as they may introduce noise during the model training process. Subsequently, we implement standard text preprocessing steps, including tokenisation, lower casing, stop words removal, stemming, and lemmatisation, to improve the efficiency and accuracy of the training process. These preprocessing techniques help reduce the complexity of the text data and allow the models to focus on relevant features that better represent the characters' personalities.

7.4.2.3 Model Training

The personality typologies suggest that a character or an individual's personality can be defined by combining the four traits for the MBTI and the five traits for the Big Five. As mentioned in section 7.4.1.1, we consider each personality trait to be dichotomous, meaning that it contains only two possible categories to which an individual or a character may belong. Consequently, our objective is to generate classification models that facilitate the identification of a character's category for each personality trait. We propose comparing several supervised learning approaches that produce binary classifiers to achieve this.

We employ five machine learning algorithms for text classification and one convolutional neural network to generate various classifiers for each personality trait. We follow the 10-fold cross-validation to thoroughly evaluate and compare the performance of the different algorithms in classifying characters based on their respective personality traits.

7.4.3 Supervised Machine Learning Algorithms

The academic literature demonstrates that logistic regression algorithms (LR), support vector machine algorithms (SVM), decision tree algorithms (DT), Naive Bayes algorithms (NB), and random forest algorithms (RF) are well-established and effective supervised machine learning techniques for natural language processing and text classification tasks [242].

Moreover, to enable a comprehensive comparison, we employ two-word embedding methods to capture the semantics and similarity of words in our characters' documents. On the one hand, we utilise a bag-of-words algorithm to generate vector representations for words by considering each word count in a document as a feature. On the other hand, we employ pre-trained word vectors, specifically GloVe [243], to represent words in the documents. In essence, we train eighteen binary classifiers for each personality trait. The bag-of-words method generally yields better results than using the pre-trained vectors. Consequently, our analysis only presents the training results obtained using the bag-of-words method.

7.4.4 Logistic Regression Algorithm

One prevalent approach to text-based personality classification is logistic regression (LR). LR employs a logistic function to determine the probability of a categorical dependent variable. It is commonly used for binary problems, where LR compares the probability with a predefined threshold to assign a binary label to a document. As shown in Table 7.5, the overall accuracy and F1-Score are relatively low, rendering LR unsuitable as a classifier for categorising characters' documents among the sixteen MBTI classes. This suggests that alternative methods may be necessary to achieve more accurate and reliable classification results for this particular task.

TABLE 7.5: Logistic Regression Classifiers Evaluation

MBTI Classifiers	Accuracy	F1-Score
E/I	0.714	0.718
S/N	0.700	0.589
T/F	0.715	0.752
J/P	0.734	0.780
Overall	0.262	0.248
Average	0.716	0.710
Big Five Classifiers	Accuracy	F1-Score
EXT	0.691	0.674
OPN	0.708	0.774
AGR	0.743	0.610
CON	0.739	0.739
NEU	0.728	0.767
Overall	0.196	0.180
Average	0.722	0.713

7.4.5 Support Vector Machine Algorithm

The Support Vector Machine (SVM) algorithm is a method that facilitates binary classification by utilising a hyperplane as a boundary and maximising the distance

between classes to distinguish between positive and negative instances. The evaluation of the models generated using SVM is presented in Table 7.6. This table illustrates the performance of SVM in our context.

TABLE 7.6: Support Vector Machine Classifiers Evaluation

MBTI Classifiers	Accuracy	F1-Score
E/I	0.677	0.687
S/N	0.665	0.560
T/F	0.674	0.714
J/P	0.710	0.752
Overall	0.215	0.207
Average	0.682	0.678
Big Five Classifiers	Accuracy	F1-Score
EXT	0.668	0.645
OPN	0.675	0.742
AGR	0.736	0.623
CON	0.686	0.688
NEU	0.700	0.741
Overall	0.159	0.151
Average	0.693	0.687

7.4.6 Naive Bayes Algorithm

Another text classification method is Naive Bayes, which follows Bayes' Theorem to calculate conditional probability. Naive Bayesian algorithms rely on prior knowledge and probability, assuming that each feature is independent of the others. This algorithm is popular and effective for document classification tasks. Table 7.7 displays the evaluation results of using the Naive Bayes algorithm to generate a classifier in our specific context.

TABLE 7.7: Naive Bayes Classifiers Evaluation

MBTI Classifiers	Accuracy	F1-Score
E/I	0.618	0.576
S/N	0.681	0.542
T/F	0.683	0.725
J/P	0.683	0.755
Overall	0.196	0.171
Average	0.666	0.650
Big Five Classifiers	Accuracy	F1-Score
EXT	0.632	0.667
OPN	0.656	0.698
AGR	0.743	0.610
CON	0.676	0.623
NEU	0.606	0.588
Overall	0.126	0.104
Average	0.663	0.637

7.4.7 Decision Tree Algorithm

The decision tree algorithm partitions the population into homogeneous subsets by hierarchically utilising the most important features. This approach facilitates the binary classification required for this study. The evaluation results of the generated classifiers using the decision tree algorithm are presented in Table 7.8.

TABLE 7.8: Decision Tree Classifiers Evaluation

MBTI Classifier	Accuracy	F1-Score
E/I	0.618	0.624
S/N	0.559	0.450
T/F	0.599	0.648
J/P	0.633	0.683
Overall	0.131	0.124
Average	0.602	0.601
Big Five Classifiers	Accuracy	F1-Score
EXT	0.613	0.599
OPN	0.639	0.715
AGR	0.633	0.478
CON	0.641	0.625
NEU	0.632	0.675
Overall	0.100	0.086
Average	0.632	0.617

7.4.8 Random Forest Algorithm

Lastly, we employ the random forest (RF) algorithm to generate the classifiers. As the name implies, RF is an ensemble of decision trees, referred to as a forest, that classifies a new object based on its attributes. RF offers better generalisation than single decision tree algorithms by classifying each tree and taking the majority vote from the trees for the specific class. The performance of the random forest algorithm in classifying characters' documents based on their personality traits can be observed in Table 7.9.

TABLE 7.9: Random Forest Classifiers Evaluation

MBTI Classifier	Accuracy	F1-Score
E/I	0.698	0.670
S/N	0.671	0.478
T/F	0.690	0.749
J/P	0.724	0.795
Overall	0.234	0.190
Average	0.696	0.673
Big Five Classifiers	Accuracy	F1-Score
EXT	0.682	0.649
OPN	0.706	0.796
AGR	0.681	0.353
CON	0.729	0.709
NEU	0.727	0.772
Overall	0.174	0.100
Average	0.705	0.656

7.4.9 Deep Learning Algorithm Using Convolutional Neural Network

A convolutional neural network (CNN) is a type of neural network that incorporates convolutional layers, which are effective for classifying text-based documents. The classification approach using a CNN for NLP tasks, such as binary classification of documents or sentences, is described by Zhang and Wallace [244] and is roughly outlined in Figure 7.3. Initially, an embedding process is applied to represent the text using vectors. Subsequently, the documents or sentences are encoded into matrices while considering the document context and utilising the CNN architecture. Then, an attend and predict phase allows for inferring the classification using a multi-layer perceptron. In contrast to the bag-of-words method employed for our machine learning classifiers, the spaCy library relies on a bloom filter [245] for its vector representation.

Following Rahman et al.'s approach [236], a convolutional neural network (CNN) has demonstrated its effectiveness in text-based personality recognition. Moreover,

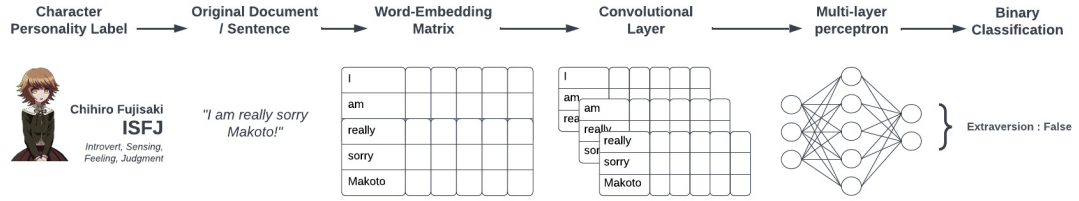


FIGURE 7.3: Classification process using CNN approach with spaCy Library

CNN enables training a multi-class classifier for the sixteen MBTI personality types. In addition to training binary classifiers for personality traits, we train a 16-class classifier to analyse whether a deep network could yield better results than our binary classifiers. The results display an accuracy value of 0.737 and an F1-Score value of 0.722 for the 16-class classifier. On the other hand, Table 7.10 presents our best results for the CNN binary classifiers after tuning multiple parameters and training the CNN multiple times with random initialisations.

TABLE 7.10: Convolutional Neural Network Classifiers Evaluation

MBTI Classifier	Accuracy	F1-Score
E/I	0.730	0.710
S/N	0.692	0.621
T/F	0.719	0.672
J/P	0.728	0.672
Overall	0,264	0,199
Average	0,717	0,669
Big Five Classifiers	Accuracy	F1-Score
EXT	0.707	0.716
OPN	0.716	0.620
AGR	0.769	0.819
CON	0.683	0.669
NEU	0.728	0.683
Overall	0,194	0,166
Average	0,721	0,701

7.4.10 Results and Discussion

7.4.10.1 Supervised Learning Methods Comparison

Table 7.11 and Table 7.12 reveal our best-performing models for the MBTI and Big Five typologies with O.Acc, O.F1, A.Acc, and A.F1 representing the overall accuracy, the overall F1-score, the average accuracy, and the average F1-score respectively.

TABLE 7.11: Methods Comparaison for the MBTI Typology

Model Type	O.Acc	O.F1	A.Acc	A.F1
Logistic regression	0,262	0,248	0.716	0,710
Support Vector Machine	0,215	0,207	0.682	0,678
Naive Bayes	0,196	0,171	0.666	0,650
Decision Tree	0,131	0,124	0.602	0.601
Random Forest	0,234	0.190	0.696	0.673
Binary CNN	0.264	0,199	0.717	0,669
16-Class CNN	0.737	0.722		

TABLE 7.12: Methods Comparaison for the Big Five Typology

Model Type	O.Acc	O.F1	A.Acc	A.F1
Logistic regression	0,196	0,180	0,722	0,713
Support Vector Machine	0,159	0,151	0,693	0,687
Naive Bayes	0,126	0,104	0,663	0,637
Decision Tree	0,100	0,086	0,632	0,617
Random Forest	0,174	0,100	0,705	0,656
Binary CNN	0,194	0,166	0,721	0,701

The results emphasise that the 16-class classifier generated by the CNN approach for the MBTI typology surpasses the overall results of the binary classifiers. In other words, the 16-Class CNN model is the preferred choice for classifying documents according to MBTI's personality types. However, game stakeholders and writers might be more interested in recognising a single personality trait than

the entire MBTI type. Although the overall accuracy and F1-score are not overwhelmingly high, the average accuracy and F1-score demonstrate better results in classifying MBTI's dichotomies and the Big Five traits. The logistic regression approach stands out among the others for both typologies, with average F1-scores of 0.710 (MBTI) and 0.713 (Big Five).

These results are consistent with the scientific literature that conducted their studies using social media posts or various personality Kaggle datasets. In line with our hypothesis question, our models exhibit promising preliminary results in predicting NPCs' personalities based on their dialogues and speeches. Furthermore, three of the six supervised learning approaches achieved better F1-Score results, as shown in Table 7.13 that compared the top classifiers by traits. In response to our research question, logistic regression, random forest, and convolutional neural network emerge as the top three approaches to classify characters' personalities among the MBTI and Big Five traits.

TABLE 7.13: Top classifiers by traits

Personality Trait	Model Type	Accuracy	F1-Score
E/I	LR	0.714	0.718
S/N	CNN	0.692	0.621
T/F	LR	0.715	0.752
J/P	RF	0.724	0.795
EXT	CNN	0.707	0.716
OPN	RF	0.706	0.796
AGR	CNN	0.769	0.819
CON	LR	0.739	0.739
NEU	RF	0.727	0.772

7.4.10.2 Limits of the study

The positive outcomes of utilising machine learning and deep learning techniques to identify non-playable characters (NPCs) personalities are evident. However, certain limitations must be addressed in future research endeavours.

Firstly, a noticeable range value of 0.198 between the highest F1-score in Table 7.13 indicates unequal performance among classifiers in detecting various personality traits. This discrepancy may be attributed to the varying distribution of dichotomies, as discussed in subsection 7.4.2.1. Moreover, the number of personality indicators like emotions in character dialogues may differ, making some traits more challenging to analyse through natural language processing (NLP) than others.

Secondly, our training utilised a dataset similar in size to those found in academic literature. However, a larger, more comprehensive dataset could lead to improved results, enhancing the reliability and validity of the generated classifiers.

Thirdly, the current findings may not be entirely neutral since the test dataset documents might share similarities with the training dataset documents. While no document appears in both datasets, they could have semantic similarities due to originating from the same NPC. We divided the entire game speeches and characters' dialogues into multiple documents to construct our complete dataset.

Lastly, as an initial investigation, we did not prioritise optimising the parameters or hyperparameters of the supervised learning algorithms. Optimising these factors could improve classifier performance, as suggested by [237, 246]. Future studies should consider these limitations and focus on refining the techniques and datasets to further enhance the detection of NPC personalities using machine learning and deep learning approaches.

7.4.11 Conclusion

This study presents initial evidence supporting the feasibility of employing machine learning and deep learning techniques to identify non-playable characters' personalities in visual novel games. We trained multiple binary classifiers and one 16-class classifier using a natural language processing approach, relying on characters' speeches and personalities. We also identified the most effective model for each trait in the MBTI and Big Five personality frameworks. This comparison has been the subject of a peer-reviewed scientific paper [247]

Despite the promising nature of these models for predicting visual novel character personalities, additional research is required to address this study's limitations.

An important question arising from this research pertains to the practical application of these models in our specific context. We applied the models to three GSJT developments. Based on the results, we revised certain dialogues to portray the characters' personalities better. Similarly, the company's writers can use our models to ensure consistency between their characters' dialogues and the intended personalities, aligning with the team employees' characteristics and stakeholders' objectives.

Moreover, natural language processing can be employed in alternative ways to improve visual novel creation processes and gameplay. Future research could explore text suggestions for writers and auto-generated dialogues based on player interactions, taking into account the sponsor company's goals. Such extensions of the current study can contribute to developing more immersive and engaging visual novel experiences while offering valuable insights into the practical applications of machine learning and deep learning for character personality identification.

Chapter Overview

In this Chapter, we delve into the application of machine learning and deep learning models for creating realistic non-playable characters (NPCs). The initial part of the chapter underscores the importance of having realistic characters in the context of soft skill recognition and interactive narrative games. Hence, in alignment with **research objective B**, we move towards the proposal of dialogue personality classifiers to aid the writing team members of a GSJT creation in the second part.

This chapter details our experiment on developing NPC speech binary personality classifiers. We focus on **MBTI and Big Five** typologies, while the tools we use for this experiment and our game pool dataset are thoroughly explained in the chapter's fourth part.

Our study underscores the efficacy of **the 16-Class Convolutional Neural Network (CNN) model** for classifying NPC personalities **within the Myers-Briggs Type Indicator (MBTI) framework**, outperforming binary classifiers. For stakeholders focused on isolating single traits, the study highlights promise for **individual trait recognition within both MBTI and the Big Five frameworks**. Notably, logistic regression and random forest emerged as consistently strong performers across both typologies. These findings are corroborated by existing scientific literature and suggest that **logistic regression, random forest, and CNN** are the leading methodologies for personality classification.

Our study highlights the promise of machine learning in characterising NPC personalities but also notes key limitations. These include **potential bias in the dataset, and lack of algorithmic optimisation**. Future work should address these issues to improve model reliability and effectiveness.

Chapter 8

General Conclusion

8.1 Our Contribution

This research was embarked upon with the goal of investigating the intersection of game-based approaches, psychometrics, and soft skills recognition. Over the course of this research, we investigated a wide range of topics, including the nature of soft skills, their importance in today's world, and the challenges that arise when trying to measure them. This broad exploration allowed us to gain a deep understanding of the many facets of our research area.

Our work began by delving into the history of game-based approaches, including their origins and their evolution, notably, the shift from purely entertainment-oriented games to serious games, edutainment games, and the emerging phenomenon of gamification and toyification. This exploration was critical in forming a foundation upon which we built the rest of our work.

A detailed analysis of related works in the realm of Human Resources and game-based approaches was performed. We focused particularly on the use of game-based approaches in recruitment, selection, and candidate experience enhancement. We also explored how these approaches address bias and objectivity and how they play a role in expanding the candidate pool.

Building upon this, we examined the use of game-based soft skill assessment methods, mainly focusing on the use of simulations, gamified situational judgment tests,

and interactive narrative games in recruitment and selection. A detailed analysis of the stakeholders, design, and conception of these tests was also conducted, steering us towards the delineation of our research objective's scope.

We promptly recognized two fundamental aspects. Firstly, the requirement for effective methodologies and approaches to foster seamless cohesion and collaboration among stakeholders engaged in GSJT creation. Secondly, a comprehensive analysis was carried out on the stakeholders, test design, and conception, which significantly influenced the development of tailored tools and approaches. These resources were specifically aimed at assisting stakeholders in achieving their inherent objectives efficiently.

Given the relative immaturity of GSJT research, we extended our review to encompass psychometric approaches and video game design methods with specificity for interactive narrative games. These comprehensive studies allowed us to analyze the strategies used and validated by experiences in these various domains. We endeavoured to integrate the best elements from each field into a list of fundamental characteristics for building a GSJT design and production methodology and structuring a game-based approach model for recruitment and selection purposes.

To address the first requirement for a cohesive approach to generate a GSJT, we started by proposing a preliminary framework that detailed the stages involved in developing a GSJT and then moved on to a more refined modular model conjugated with a drama manager to guide GSJT designers. For each phase of the framework, we detailed the roles and responsibilities of the involved stakeholders. This drama manager also assumes a primary function in facilitating the replayability of the gamified assessment tools, which frequently face challenges in interactive narrative games. The aim is to guide GSJT designers in focusing on the evaluation's reliability and validity in the context of the GSJT rather than on the game's narrative and storyline, which can often become intricate in Interactive Narrative Games.

We also designed a computational psychometric scoring system to aid the psychometric teams in a hybrid approach to score a GSJT, which was a significant step forward in streamlining the scoring process and maintaining the objectivity of the assessments. This system was formulated keeping in mind the challenges and limitations of existing scoring methods and is expected to enhance the efficiency and effectiveness of GSJT assessments. We validated our theories through carefully

designed experiments, and the results affirmed preliminary findings in the efficacy of our proposed frameworks and hybrid scoring systems.

In the final chapter of this work, we underscored the importance of realistic characters for the recognition of soft skills as a fundamental element for accurate and dependable soft skill evaluation. We introduced dialogue personality classifiers aimed at aiding narrative writers in enhancing the immersive quality and realism of NPCs. The process for developing binary classifiers for NPC dialogues is comprehensively explained, encompassing stages like data collection, labelization, preprocessing, and model training. We conducted extensive experiments utilizing a range of machine learning and deep learning algorithms, including Logistic Regression, Support Vector Machine, Naive Bayes, Decision Trees, Random Forest, and Convolutional Neural Networks. Our in-depth analysis of the results offers vital insights into the capabilities of machine learning and deep learning as tools to assist GSJT designers in NPC creation.

In conclusion, this research represents one of the first steps forward in our understanding and application of game-based approaches in the field of soft skills recognition. This research provides a comprehensive guide to designing, implementing, and scoring GSJTs, a tool that is anticipated to be used by the human resources industry by offering a novel, engaging, and effective method for soft skill and behavioural profile recognition. However, like any scientific endeavour, this work raises new questions and paves the way for future research. We expect further refinements to our frameworks and scoring system as they are applied in various contexts and as technology continues to advance.

8.2 Perspectives and Future Research

Our thesis operates within a framework of specific boundaries and constraints that we delimited in chapter 1, which are not only inherent but also instrumental in guiding our research process. It is within these confines that our study derives its focus, allowing for a targeted and in-depth exploration of the chosen subject matter. As with all research endeavours, this thesis, while initiating in its approach to melding game-based approaches, psychometrics, and soft skill recognition, is no exception to the presence of certain limitations. Acknowledging these constraints not only ensures transparency and candour in our work but also provides a roadmap

for future researchers. We've included discussion and limitations for each proposal in the individual chapters. Therefore, this section provides an overview of the thesis's general constraints, aiming to clarify the context for interpreting our findings and to guide future research.

- **Scope and Generalization:** The universal applicability of game-based techniques. and specifically, GSJT remains a significant question within the academic community. Variations in sectorial demands, cultural values, and regional norms create a multifaceted landscape of soft skills recognition. Our study and proposals may represent a snapshot of this vast terrain. Different sectors, from technology to healthcare to the arts, present unique sets of challenges or opportunities. For instance, the healthcare sector may prioritize empathy and communication skills, while the technology sector might lean towards collaboration and adaptability. Additionally, cultural contexts play a pivotal role in shaping the definitions and perceptions of 'valued' soft skills. For instance, certain cultures may prioritize harmony and group cohesion, while others emphasize individual innovation. As such, future research endeavours could undertake a multi-pronged approach, focusing on both sectorial and cultural variations. Cross-sectional studies that juxtapose the efficacy and adaptability of GSJT techniques across these diverse terrains would significantly enhance our understanding.
- **Stakeholder Analysis Specificity:** A well-rounded research perspective necessitates a comprehensive understanding of various stakeholder views, especially when it involves subjective domains such as soft skills recognition. While our study has tapped into the perspectives of key stakeholder groups, there exist stakeholders on the periphery that might not have been adequately captured. These stakeholders, especially from underrepresented or marginalized communities, could offer alternative insights, challenging the predominant narratives. For instance, stakeholders from non-traditional educational or HR backgrounds, differently-abled individuals, or those from socio-economically challenged environments might have unique views on game-based recognition of soft skills. Future studies could adopt a more diversified sampling strategy, ensuring representation from varied demographics. Leveraging mixed-method approaches, combining qualitative interviews with large-scale surveys, could further enhance the depth and breadth of stakeholder analysis.

- **Navigating the Nascent Terrain of GSJT:** With the GSJT domain still in its budding stage, our foundational understanding is continually evolving, with potential blind spots that might emerge over time. While our thesis has strived to provide a comprehensive overview of the GSJT domain, it operates within the constraints of current knowledge and this thesis's constraints. As the field matures, new paradigms might emerge, challenging our foundational assumptions. For instance, user behaviour within game environments might evolve, or new technological advancements could redefine game dynamics to enhance the reliability and validity of the assessment. Hence, a longitudinal approach that tracks GSJT's evolution could provide invaluable insights. Such an approach would not only chart the trajectory of GSJT's evolution but also identify pivotal milestones, challenges, and shifts in the domain.
- **Iterative Evolution of Frameworks:** The proposed preliminary GSJT creation frameworks described in chapters 4 and 5, developed on present understandings and best practices, represent an initial blueprint. The fluid and dynamic nature of research domains, especially those intersecting technology and human behaviour, necessitates adaptable and evolving frameworks. The GSJT creation framework, while comprehensive in its current form, stands on the knowledge and best practices of today. As the domain interacts with real-world scenarios, certain elements of the framework might demand recalibration. For instance, feedback from users, HR professionals, or game designers might point towards elements that need refinement or even a complete overhaul. Action research methodologies, wherein researchers collaborate with practitioners in real-world settings, could be invaluable. This would allow for continuous feedback loops, with iterative modifications to the GSJT framework based on real-world insights and challenges.
- **Balancing Gameplay and Assessment:** Prioritizing replayability using the drama manager might introduce challenges. Striking a balance between creating an immersive gaming experience and ensuring accurate skill assessment remains one of the most formidable challenges in game-based soft skills recognition. The risk of favouring immersive gameplay over accurate assessment or vice-versa presents potential pitfalls. For example, an overly engaging game narrative might distract from the primary goal of skills assessment. Conversely, a game too focused on assessment might compromise

user engagement, reducing its effectiveness. Future research could design experimental setups that create variations in game narratives and assessment focus, subsequently evaluating the efficacy of each variation. Such experiments would help in isolating potential trade-offs, leading to strategies that ensure an optimal balance between immersion and assessment.

- **Computational Psychometric Scoring:** The innovative computational scoring system is a stride forward, merging psychometrics with computational advances. Yet, as computational techniques evolve and our understanding of psychometrics deepens, there's a constant need to refine and redefine our scoring systems. The existing computational scoring system, while innovative, operates within the boundaries of current algorithms and psychometric principles. With rapid advancements in AI and evolving nuances in psychometric understanding, future challenges or even paradigm shifts are anticipated. Meta-analytic studies could be useful, comparing the efficacy of various computational scoring systems across diverse game-based platforms and populations. Furthermore, exploring collaboration between AI experts and psychometricians could lead to the creation of more sophisticated, accurate, and universally applicable scoring algorithms.
- **Enhancing NPC Realism and navigating the complexities of Machine Learning:** The deployment of machine learning and deep learning for NPC enhancement brings about its set of challenges. Beyond the constraints of the algorithms themselves, the quality and breadth of training data, potential algorithmic biases, and the transient nature of AI research might lead to newer methodologies that challenge or surpass current implementations. In addition, the complexities of human interactions, rife with subtle cues and nuanced responses, present challenges when replicated in virtual environments through NPCs. Achieving high degrees of realism in NPCs is pivotal for effective soft skill recognition, especially when assessing interpersonal dynamics. The current generation of NPCs might not fully capture the subtleties of human behaviour. For instance, nuances like sarcasm, humour, or cultural idioms might be challenging to replicate authentically. Interdisciplinary research combining insights from behavioural psychology, human-computer interaction, narrative design, and even fields like sociology or anthropology could pave the way for more sophisticated and nuanced NPC designs. Such collaboration could delve deeper into understanding human

interactions and subsequently design NPCs that mirror these complexities accurately.

- **From Theory to Practice: Operational Feasibility:** The transition from research-based frameworks to real-world applications often unveils unforeseen challenges, underscoring the need for practical feasibility studies. Scalability, cost implications, user receptivity, and adaptability within organisational structures are just a few of the myriad challenges that might arise when operationalizing the proposed frameworks. Pilot studies, wherein the frameworks are introduced on a smaller scale within real-world HR settings, could provide invaluable insights. Detailed post-implementation reviews could shed light on challenges faced, areas of improvement, and the overall efficacy of the framework in real-world scenarios. Collaborative research involving academia, industry professionals, and end-users could ensure that the research outcomes are both theoretically robust and practically applicable.

In conclusion, while this thesis has endeavoured to chart new territories within the domain of game-based soft skills recognition through gamified situational judgment tests, the vast landscape of potential research beckons subsequent explorations. The extended discussions above aim to provide a roadmap for researchers, practitioners, and stakeholders, illuminating the various avenues that remain to be explored, challenged, and refined.

Bibliography

- [1] Kristin Stoeffler, Yigal Rosen, Maria Bolsinova, and Alina A von Davier. Gamified assessment of collaborative skills with chatbots. In *Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part II 19*, pages 343–347. Springer, 2018.
- [2] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining” gamification”. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15, 2011.
- [3] David R Michael and Sandra L Chen. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.
- [4] Thomas M Connolly, Elizabeth A Boyle, Ewan MacArthur, Thomas Hainey, and James M Boyle. A systematic literature review of empirical evidence on computer games and serious games. *Computers & education*, 59(2):661–686, 2012.
- [5] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work?—a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences*, pages 3025–3034. Ieee, 2014.
- [6] Katie Seaborn and Deborah I Fels. Gamification in theory and action: A survey. *International Journal of human-computer studies*, 74:14–31, 2015.
- [7] Miriam L Matteson, Lorien Anderson, and Cynthia Boyden. ” soft skills”: A phrase in search of meaning. *portal: Libraries and the Academy*, 16(1): 71–88, 2016.
- [8] Marcel M Robles. Executive perceptions of the top 10 soft skills needed in today’s workplace. *Business communication quarterly*, 75(4):453–465, 2012.

- [9] James J Heckman and Tim Kautz. Hard evidence on soft skills. *Labour economics*, 19(4):451–464, 2012.
- [10] Murray R Barrick and Michael K Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26, 1991.
- [11] Wendy L Bedwell, Stephen M Fiore, and Eduardo Salas. Developing the future workforce: An approach for integrating interpersonal skills into the mba classroom. *Academy of Management Learning & Education*, 13(2):171–186, 2014.
- [12] Angela Duckworth and James J Gross. Self-control and grit: Related but separable determinants of success. *Current directions in psychological science*, 23(5):319–325, 2014.
- [13] Shaheen Majid, Zhang Liming, Shen Tong, and Siti Raihana. Importance of soft skills for education and career success. *International Journal for Cross-Disciplinary Subjects in Education*, 2(2):1037–1042, 2012.
- [14] Jane Andrews and Helen Higson. Graduate employability, ‘soft skills’ versus ‘hard’ business knowledge: A european study. *Higher education in Europe*, 33(4):411–422, 2008.
- [15] Rosli Ibrahim, Ali Boerhannoeddin, and Kazeem Kayode Bakare. The effect of soft skills and training methodology on employee performance. *European Journal of Training and Development*, 2017.
- [16] Daniel Goleman. Emotional intelligence. why it can matter more than iq. *Learning*, 24(6):49–50, 1996.
- [17] James Manyika and Kevin Sneader. Ai, automation, and the future of work: Ten things to solve for. 2018.
- [18] Allen I Huffcutt and Winfred Arthur. Hunter and hunter (1984) revisited: Interview validity for entry-level jobs. 1994.
- [19] James A Breaugh. Employee recruitment: Current knowledge and important areas for future research. *Human Resource Management Review*, 18(3):103–118, 2008.

- [20] M Susan Taylor and Christopher J Collins. Organizational recruitment: Enhancing the intersection of research and practice. 2000.
- [21] Robert Gatewood, Hubert S Feild, and Murray Barrick. *Human resource selection*. Cengage Learning, 2015.
- [22] Wayne F Cascio and Herman Aguinis. Applied psychology in human resource management. 2005.
- [23] Frank L Schmidt and John E Hunter. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124(2):262, 1998.
- [24] Barbara Cimatti. Definition, development, assessment of soft skills and their role for the quality of organizations and enterprises. *International Journal for quality research*, 10(1):97, 2016.
- [25] Sowmya Rasipuram and Dinesh Babu Jayagopi. Automatic multimodal assessment of soft skills in social interactions: a review. *Multimedia Tools and Applications*, 79(19-20):13037–13060, 2020.
- [26] Angela L Duckworth and David Scott Yeager. Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational researcher*, 44(4):237–251, 2015.
- [27] Stewart I Donaldson and Elisa J Grant-Vallone. Understanding self-report bias in organizational behavior research. *Journal of business and Psychology*, 17:245–260, 2002.
- [28] Delroy L Paulhus, Simine Vazire, et al. The self-report method. *Handbook of research methods in personality psychology*, 1(2007):224–239, 2007.
- [29] Brian S Connelly and Deniz S Ones. An other perspective on personality: meta-analytic integration of observers’ accuracy and predictive validity. *Psychological bulletin*, 136(6):1092, 2010.
- [30] Gary J Summers. Today’s business simulation industry. *Simulation & Gaming*, 35(2):208–241, 2004.
- [31] Michael A McDaniel, Nathan S Hartman, Deborah L Whetzel, and W LEE GRUBB III. Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel psychology*, 60(1):63–91, 2007.

- [32] Jesus F Salgado. The five factor model of personality and job performance in the european community. *Journal of Applied psychology*, 82(1):30, 1997.
- [33] Jennifer Herrity. 8 personality tests used in psychology (and by employers), 2022. URL <https://www.indeed.com/career-advice/career-development/types-of-personality-test>.
- [34] Paul T Costa Jr and Robert R McCrae. *The Revised Neo Personality Inventory (neo-pi-r)*. Sage Publications, Inc, 2008.
- [35] Robert R McCrae and Paul T Costa Jr. Personality trait structure as a human universal. *American psychologist*, 52(5):509, 1997.
- [36] Jack Block. A contrarian view of the five-factor approach to personality description. *Psychological bulletin*, 117(2):187, 1995.
- [37] Dan P McAdams. The five-factor model in personality: A critical appraisal. *Journal of personality*, 60(2):329–361, 1992.
- [38] Isabel Briggs Myers, Mary H McCaulley, and Robert Most. Manual: A guide to the development and use of the myers-briggs type indicator. (*No Title*), 1985.
- [39] Thomas Moore. Personality tests are back. *Fortune*, 115(7):74, 1987.
- [40] David J Pittenger. The utility of the myers-briggs type indicator. *Review of educational research*, 63(4):467–488, 1993.
- [41] Anthony J Devito. Review of myers-briggs type indicator. *The ninth mental measurements yearbook*, 2:1030–1032, 1985.
- [42] Adrian Furnham. The big five versus the big four: the relationship between the myers-briggs type indicator (mbti) and neo-pi five factor model of personality. *Personality and individual differences*, 21(2):303–307, 1996.
- [43] Tammy L Bess and Robert J Harvey. Bimodal score distributions and the myers-briggs type indicator: fact or artifact? *Journal of personality assessment*, 78(1):176–186, 2002.
- [44] Michael C Ashton and Kibeom Lee. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and social psychology review*, 11(2):150–166, 2007.

- [45] Iza Marfisi-Schottman. *Méthodologie, modèles et outils pour la conception de Learning Games*. PhD thesis, Lyon, INSA, 2012.
- [46] David Nieborg. *America's army: More than a game*, 2004.
- [47] Clark Aldrich. *The complete guide to simulations and serious games: How the most valuable content will be created in the age beyond Gutenberg to Google*. John Wiley & Sons, 2009.
- [48] J Millet. Êtes-vous sûr du type de jeu que vous avez choisi pour votre serious game? *Serious Game Expo*, 2010.
- [49] Konstantina Georgiou, Athanasios Gouras, and Ioannis Nikolaou. Gamification in employee selection: The development of a gamified assessment. *International journal of selection and assessment*, 27(2):91–103, 2019.
- [50] Michela Mortara, Chiara Eva Catalano, Francesco Bellotti, Giusy Fiucci, Minica Houry-Panchetti, and Panagiotis Petridis. Learning cultural heritage by serious games. *Journal of Cultural Heritage*, 15(3):318–325, 2014.
- [51] Valerie J Shute, Matthew Ventura, and Yoon Jeon Kim. Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, 106(6):423–430, 2013.
- [52] Sebastian Deterding. Make-believe in gameful and playful design. *Digital make-believe*, pages 101–124, 2016.
- [53] Leslie Kurke. Ancient greek board games and how to play them. *Classical philology*, 94(3):247–267, 1999.
- [54] Arthur A Krentz. Play and education in plato's republic. In *The Paideia Archive: Twentieth World Congress of Philosophy*, volume 29, pages 199–207, 1998.
- [55] Sun Tzu. The art of war. In *Strategic Studies*, pages 63–91. Routledge, 2008.
- [56] Ed Halter. *From Sun Tzu to xbox: War and video games*. Thunder's Mouth Press, 2006.
- [57] Carl Von Clausewitz. *On war*, volume 1. Jazzybee Verlag, 1950.
- [58] Richard Hodgetts. Management gaming for didactic purposes: A new look. *Simulation & Games*, 1(1):55–66, 1970.

- [59] Yvon Feaster, Farha Ali, and Jason O Hallstrom. Serious toys: teaching the binary number system. In *Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education*, pages 262–267, 2012.
- [60] Peter P Perla. *The art of wargaming: A guide for professionals and hobbyists*. Naval Institute Press, 1990.
- [61] Damien Djaouti, Julian Alvarez, Jean-Pierre Jessel, and Olivier Rampnoux. Origins of serious games. *Serious games and edutainment applications*, pages 25–43, 2011.
- [62] Clark C Abt. *Serious games*. University press of America, 1987.
- [63] Michael Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.
- [64] Ben Sawyer and David Rejeski. Serious games: Improving public policy through game-based learning and simulation. *Woodrow Wilson International Center for Scholars*, 31(1), 2002.
- [65] Julian Alvarez, Damien Djaouti, et al. An introduction to serious game definitions and concepts. *Serious games & simulation for risks management*, 11(1):11–15, 2011.
- [66] Donald R Jansiewicz. *The New Alexandria simulation: a serious game of state and local politics*. Canfield Press San Francisco, CA, 1973.
- [67] D Meadows, T Iddman, and D Shannon. Fish banks, ltd: Game administrator’s manual. *Laboratory of Interactive Learning, University of New Hampshire, Durham, USA*, 1993.
- [68] Mélanie Despeisse. Games and simulations in industrial engineering education: a review of the cognitive and affective learning outcomes. In *2018 Winter Simulation Conference (WSC)*, pages 4046–4057. IEEE, 2018.
- [69] David Crookall and Barbara Steinwachs. Book and simulation/game reviews: Index to volumes 16 through 21. *Simulation & Gaming*, 21(4):488–498, 1990.
- [70] Joey J Lee, Stephanie Pfirman, Thomas Toynton, and Eduard Matamoros. Ecochains: A multiplayer card game to teach food webs, climate change and systems thinking.

- [71] Jeff Van West and Kevin Lane-Cummings. *Microsoft flight simulator X for pilots: Real world training*. John Wiley & Sons, 2007.
- [72] Jefferson M Koonce and William J Bramble Jr. Personal computer-based flight training devices. *The international journal of aviation psychology*, 8(3):277–292, 1998.
- [73] Alfred T Lee. *Flight simulation: virtual environments in aviation*. Routledge, 2017.
- [74] Cyril Brom, Michal Preuss, and Daniel Klement. Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? a quasi-experimental study. *Computers & Education*, 57(3):1971–1988, 2011.
- [75] David Squires. Educational software for constructivist learning environments: Subversive use and volatile design. *Educational technology*, pages 48–54, 1999.
- [76] Pablo Moreno-Ger, Daniel Burgos, Iván Martínez-Ortiz, José Luis Sierra, and Baltasar Fernández-Manjón. Educational game design for online education. *Computers in Human Behavior*, 24(6):2530–2540, 2008.
- [77] Eric Klopfer and Scot Osterweil. The boom and bust and boom of educational games. *Transactions on edutainment IX*, pages 290–296, 2013.
- [78] Kurt Squire. *Game-based learning: Present and future state of the field*. Masie center e-learning consortium New York, NY, 2005.
- [79] Lee Wilson. Best practices for using games & simulations in the classroom guidelines for k 12 educators. 2009.
- [80] Amy Bruckman. Can educational be fun. In *Game developers conference*, volume 99, pages 75–79, 1999.
- [81] Carol Luckhardt Redfield, Diane L Gaither, and Neil M Redfield. Cots computer game effectiveness. In *Handbook of research on effective electronic gaming in education*, pages 277–294. IGI Global, 2009.
- [82] Linda M Raffaele Mendez, Catherine A Pelzmann, and Michael J Frank. Engaging struggling early readers to promote reading success: A pilot study of reading by design. *Reading & Writing Quarterly*, 32(3):273–297, 2016.

- [83] Alice Medalia, Hanni Dorn, and Sniezyna Watras-Gans. Treating problem-solving deficits on an acute care psychiatric inpatient unit. *Psychiatry Research*, 97(1):79–88, 2000.
- [84] Terry Carroll. ” carmen sandiego:” crime can pay when it comes to learning. instructional technology. *Social Education*, 59(3):165–69, 1995.
- [85] Richard De Lisi and Jennifer L Wolford. Improving children’s mental rotation accuracy with computer game playing. *The Journal of genetic psychology*, 163(3):272–282, 2002.
- [86] David Williamson Shaffer and James Paul Gee. *How computer games help children learn*. Springer, 2006.
- [87] James Paul Gee. What video games have to teach us about learning and literacy. *Computers in entertainment (CIE)*, 1(1):20–20, 2003.
- [88] Mark Peterson. Massively multiplayer online role-playing games as arenas for second language learning. *Computer Assisted Language Learning*, 23(5):429–439, 2010.
- [89] Yolanda A Rankin, McKenzie McNeal, Marcus W Shute, and Bruce Gooch. User centered game design: evaluating massive multiplayer online role playing games for second language acquisition. In *Proceedings of the 2008 ACM SIGGRAPH symposium on Video games*, pages 43–49, 2008.
- [90] Isara Kongmee, Rebecca Strachan, Catherine Montgomery, and Alison Pickard. Using massively multiplayer online role playing games (mmorpgs) to support second language learning: Action research in the real and virtual world. 2011.
- [91] Marcus D Childress and Ray Braswell. Using massively multiplayer online role-playing games for online learning. *Distance Education*, 27(2):187–196, 2006.
- [92] Bonnie A Nardi, Stella Ly, and Justin Harris. Learning conversations in world of warcraft. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS’07)*, pages 79–79. IEEE, 2007.
- [93] Theofylaktos Anastasiadis, Georgios Lampropoulos, and Kerstin Siakas. Digital game-based learning and serious games in education. *International*

- Journal of Advances in Scientific Research and Engineering*, 4(12):139–144, 2018.
- [94] Marc Prensky. Digital game-based learning. *Computers in Entertainment (CIE)*, 1(1):21–21, 2003.
- [95] Séverine Erhel and Eric Jamet. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & education*, 67:156–167, 2013.
- [96] Yuh-Ming Cheng, Shi-Jer Lou, Sheng-Huang Kuo, and Ru-Chu Shih. Investigating elementary school students’ technology acceptance by applying digital game-based learning to environmental education. *Australasian Journal of Educational Technology*, 29(1), 2013.
- [97] Carlo Perrotta, Gill Featherstone, Helen Aston, and Emily Houghton. Game-based learning: Latest evidence and future directions. *Slough: nfer*, pages 1–49, 2013.
- [98] Hiller A Spires. Digital game-based learning: What’s literacy got to do with it? *Journal of Adolescent & Adult Literacy*, 59(2):125–130, 2015.
- [99] Robert E Slavin. Cooperative learning. *Review of educational research*, 50(2):315–342, 1980.
- [100] Joseph S Krajcik and Phyllis C Blumenfeld. *Project-based learning*. na, 2006.
- [101] John R Savery. Overview of problem-based learning: Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, 9(2):5–15, 2015.
- [102] Sebastian Deterding. The ambiguity of games: Histories and discourses of a gameful world. 2014.
- [103] Stephen J Aguilar, Caitlin Holman, and Barry J Fishman. Game-inspired design: Empirical evidence in support of gameful learning environments. *Games and Culture*, 13(1):44–70, 2018.
- [104] Jeanne Nakamura, Mihaly Csikszentmihalyi, et al. The concept of flow. *Handbook of positive psychology*, 89:105, 2002.

- [105] Mihaly Csikszentmihalyi and Kevin Rathunde. The measurement of flow in everyday life: toward a theory of emergent motivation. 1993.
- [106] Zachary H Morford, Benjamin N Witts, Kenneth J Killingsworth, and Mark P Alavosius. Gamification: the intersection between behavior analysis and game design technologies. *The Behavior Analyst*, 37:25–40, 2014.
- [107] Roger Caillois. *Man, play, and games*. University of Illinois press, 2001.
- [108] Andrés Lucero, Evangelos Karapanos, Juha Arrasvuori, and Hannu Korhonen. Playful or gameful? creating delightful user experiences. *interactions*, 21(3):34–39, 2014.
- [109] Kurt Squire. From content to context: Videogames as designed experience. *Educational researcher*, 35(8):19–29, 2006.
- [110] David Bustard, Michaela Black, et al. Game inspired tool support for e-learning processes. *Electronic Journal of e-Learning*, 7(2):pp101–110, 2009.
- [111] Atreyi Kankanhalli, Mahdiah Taher, Huseyin Cavusoglu, and Seung Hyun Kim. Gamification: A new paradigm for online user engagement. 2012.
- [112] Richard M Ryan and Edward L Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.
- [113] Richard M Ryan, C Scott Rigby, and Andrew Przybylski. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, 30:344–360, 2006.
- [114] Ryan Rogers. The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach. *Computers in Human Behavior*, 73:446–450, 2017.
- [115] Jean-Nicolas Proulx, Margarida Romero, and Sylvester Arnab. Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning. *Simulation & Gaming*, 48(1):81–97, 2017.
- [116] Robin Hunicke, Marc LeBlanc, Robert Zubek, et al. Mda: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, volume 4, page 1722. San Jose, CA, 2004.

- [117] Katie Salen Tekinbas and Eric Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2003.
- [118] Paul Cobb and Janet Bowers. Cognitive and situated learning perspectives in theory and practice. *Educational researcher*, 28(2):4–15, 1999.
- [119] Mihaly Csikszentmihalyi, Sami Abuhamdeh, and Jeanne Nakamura. Flow. *Handbook of competence and motivation*, pages 598–608, 2005.
- [120] Ian Bogost. *Persuasive games: The expressive power of videogames*. mit Press, 2010.
- [121] Gabe Zichermann and Christopher Cunningham. *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc.", 2011.
- [122] Andrzej Marczewski. Even ninja monkeys like to play. *London: Blurb Inc*, 1(1):28, 2015.
- [123] Johan Huizinga. *Homo ludens: A study of the play-element in culture*. Routledge, 2014.
- [124] Valerie J Shute. Stealth assessment in computer-based games to support learning. *Computer games and instruction*, 55(2):503–524, 2011.
- [125] Emma Witkowski. Running from zombies. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, pages 1–8, 2013.
- [126] Simon McCallum. Gamification and serious games for personalized health. *PHealth*, 177:85–96, 2012.
- [127] Ivan L Beale, Pamela M Kato, Veronica M Marin-Bowling, Nicole Guthrie, and Steve W Cole. Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, 41(3):263–270, 2007.
- [128] Paulette Guitard, Francine Ferland, and Élisabeth Dutil. Toward a better understanding of playfulness in adults. *OTJR: Occupation, Participation and Health*, 25(1):9–22, 2005.

- [129] Cristyne Hébert and Jennifer Jenson. Teaching with sandbox games: Minecraft, game-based learning, and 21st century competencies. *Canadian Journal of Learning and Technology*, 46(3), 2020.
- [130] Per Kristiansen and Robert Rasmussen. *Building a better business using the Lego serious play method*. John Wiley & Sons, 2014.
- [131] Steve Nebel, Sascha Schneider, and Günter Daniel Rey. Mining learning and crafting scientific experiments: a literature review on the use of minecraft in education and research. *Journal of Educational Technology & Society*, 19(2):355–366, 2016.
- [132] Dave Gray, Sunni Brown, and James Macanuso. *Gamestorming: A playbook for innovators, rulebreakers, and changemakers*. ” O’Reilly Media, Inc.”, 2010.
- [133] Michael B Armstrong, Richard N Landers, and Andrew B Collmus. Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In *Emerging research and trends in gamification*, pages 140–165. IGI Global, 2016.
- [134] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. A temporally based framework and taxonomy of team processes. *Academy of management review*, 26(3):356–376, 2001.
- [135] Saman Bina, Jeffrey Mullins, and Stacie Petter. Examining game-based approaches in human resources recruitment and selection: A literature review and research agenda. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, page 1325, 2021.
- [136] Andrew B Collmus and Richard N Landers. Game-framing to improve applicant perceptions of cognitive assessments. *Journal of Personnel Psychology*, 2019.
- [137] Panagiotis Gkorezis, Konstantina Georgiou, Ioannis Nikolaou, and Anna Kyriazati. Gamified or traditional situational judgement test? a moderated mediation model of recommendation intentions via organizational attractiveness. *European Journal of Work and Organizational Psychology*, 30(2): 240–250, 2021.

- [138] Denise Marie Kuepper, Kristina Klein, and Franziska Voelckner. Gamifying employer branding: An integrating framework and research propositions for a new hr approach in the digitized economy. *Human Resource Management Review*, 31(1):100686, 2021.
- [139] Richard N Landers, Elena M Auer, and Joseph D Abraham. Gamifying a situational judgment test with immersion and control game elements: Effects on applicant reactions and construct validity. *Journal of Managerial Psychology*, 35(4):225–239, 2020.
- [140] Stephen A Woods, Sara Ahmed, Ioannis Nikolaou, Ana Cristina Costa, and Neil R Anderson. Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1):64–77, 2020.
- [141] Xinyue Ren. Stealth assessment embedded in game-based learning to measure soft skills: A critical review. *Game-Based Assessment Revisited*, pages 67–83, 2019.
- [142] Lynda Donovan and P Lead. The use of serious games in the corporate sector. *A State of the Art Report. Learnovate Centre (December 2012)*, 2012.
- [143] Daniel M Cable and Timothy A Judge. Person–organization fit, job choice decisions, and organizational entry. *Organizational behavior and human decision processes*, 67(3):294–311, 1996.
- [144] Amy L Kristof. Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel psychology*, 49(1): 1–49, 1996.
- [145] Charles A O’Reilly III, Jennifer Chatman, and David F Caldwell. People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of management journal*, 34(3):487–516, 1991.
- [146] Karen Robson, Kirk Plangger, Jan H Kietzmann, Ian McCarthy, and Leyland Pitt. Game on: Engaging customers and employees through gamification. *Business horizons*, 59(1):29–36, 2016.
- [147] Richard N Landers, Kristina N Bauer, and Rachel C Callan. Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, 71:508–515, 2017.

- [148] Richard N Landers, Michael B Armstrong, Andrew B Collmus, Salih Mujcic, and Jason Blaik. Theory-driven game-based assessment of general cognitive ability: Design theory, measurement, prediction of performance, and test fairness. *Journal of Applied Psychology*, 107(10):1655, 2022.
- [149] Iqra Obaid, Muhammad Shoaib Farooq, and Adnan Abid. Gamification for recruitment and job training: model, taxonomy, and challenges. *IEEE Access*, 8:65164–65178, 2020.
- [150] Michael B Armstrong and Richard N Landers. An evaluation of gamified training: Using narrative to improve reactions and learning. *Simulation & Gaming*, 48(4):513–538, 2017.
- [151] Richard N Landers, Elena M Auer, Andrew B Collmus, and Michael B Armstrong. Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming*, 49(3):315–337, 2018.
- [152] Franziska Leutner, Sonia-Cristina Codreanu, Suzanne Brink, and Theodoros Bitsakis. Game based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience. *Frontiers in Psychology*, 13, 2023.
- [153] Konstantina Georgiou. Can explanations improve applicant reactions towards gamified assessment methods? *International Journal of Selection and Assessment*, 29(2):253–268, 2021.
- [154] Bradford S Bell and Steve WJ Kozlowski. Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, 55(2):267–306, 2002.
- [155] Slavica Nenadić. *Gamification in HR; applicability and its importance in recruitment and selection*. PhD thesis, RIT Croatia, 2019.
- [156] Simon Trang and Welf H Weiger. The perils of gamification: Does engaging with gamified services increase users’ willingness to disclose personal information? *Computers in Human Behavior*, 116:106644, 2021.
- [157] Jack Buckley, Laura Colosimo, Rebecca Kantar, Marty McCall, and Erica Snow. Game-based assessment for education. *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, page 195, 2021.

- [158] Benjamin Schneider. The people make the place. *Personnel psychology*, 40 (3):437–453, 1987.
- [159] Oihab Allal-Chérif and Marc Bidan. Collaborative open training with serious games: Relations, culture, knowledge, innovation, and desire. *Journal of Innovation & Knowledge*, 2(1):31–38, 2017.
- [160] Manu Melwin Joy and J Assistant. An investigation into gamification as a tool for enhancing recruitment process. *Ideal Research*, 3(1):56–65, 2017.
- [161] Monika Hamori. Job-hopping to the top and other career fallacies. *Harvard business review*, 88(7-8):154–157, 2010.
- [162] D Pranaya. Job-hopping—an analytical review. *International Journal of Research in Business Management*, 2(4):67–72, 2014.
- [163] Ioannis Nikolaou. What is the role of technology in recruitment and selection? *The Spanish journal of psychology*, 24:e2, 2021.
- [164] John P Hausknecht, David V Day, and Scott C Thomas. Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel psychology*, 57(3):639–683, 2004.
- [165] Jason A Colquitt, Donald E Conlon, Michael J Wesson, Christopher OLH Porter, and K Yee Ng. Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of applied psychology*, 86 (3):425, 2001.
- [166] Talya N Bauer, Donald M Truxillo, Rudolph J Sanchez, Jane M Craig, Philip Ferrara, and Michael A Campion. Applicant reactions to selection: Development of the selection procedural justice scale (spjs). *Personnel psychology*, 54(2):387–419, 2001.
- [167] Stephen W Gilliland. The perceived fairness of selection systems: An organizational justice perspective. *Academy of management review*, 18(4):694–734, 1993.
- [168] Leah Joyce Ellison, Tara McClure Johnson, David Tomczak, Alina Siemsen, and Manuel Francisco Gonzalez. Game on! exploring reactions to game-based selection assessments. *Journal of Managerial Psychology*, 35(4):241–254, 2020.

- [169] Talya N Bauer, Donald M Truxillo, Kyle Mack, and Ana B Costa. Applicant reactions to technology-based selection: what we know so far. *Technology-Enhanced Assessment of Talent*, pages 190–223, 2011.
- [170] Richard N Landers and Michael B Armstrong. Enhancing instructional outcomes with gamification: An empirical test of the technology-enhanced training effectiveness model. *Computers in human behavior*, 71:499–507, 2017.
- [171] Talya N Bauer, Julie McCarthy, Neil Anderson, Donald M Truxillo, and Jesus F Salgado. What we know about applicant reactions to selection: Research summary and best practices. *Society for Human Resource Management and Society for Industrial and Organizational Psychology. Žiūrētā lapkričio*, 20, 2012.
- [172] Scott A Birkeland, Todd M Manson, Jennifer L Kisamore, Michael T Brannick, and Mark A Smith. A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14(4):317–335, 2006.
- [173] Richard N Landers and Diana R Sanchez. Game-based, gamified, and gamefully designed assessments for employee selection: Definitions, distinctions, design, and validation. *International Journal of Selection and Assessment*, 30(1):1–13, 2022.
- [174] Ashutosh Bishnu Murti. Why soft skills matter. *IUP Journal of Soft Skills*, 8(3), 2014.
- [175] Ann Marie Ryan and Robert E Ployhart. A century of selection. *Annual review of psychology*, 65:693–717, 2014.
- [176] Mahzarin R Banaji and Anthony G Greenwald. *Blindspot: Hidden biases of good people*. Bantam, 2016.
- [177] Hannah-Hanh D Nguyen and Ann Marie Ryan. Does stereotype threat affect test performance of minorities and women? a meta-analysis of experimental evidence. *Journal of applied psychology*, 93(6):1314, 2008.
- [178] Darina Dicheva, Christo Dichev, Gennady Agre, and Galia Angelova. Gamification in education: A systematic mapping study. *Journal of educational technology & society*, 18(3):75–88, 2015.

- [179] Jiayi Lv, Zhaoyang Sun, Hao Li, and Yubo Hou. The role of negative perfectionism and the relationship between critical thinking and the halo effect: Insights from corporate managers in human resources. *Behavioral Sciences*, 13(7):533, 2023.
- [180] Richard E Nisbett and Timothy D Wilson. The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, 35(4):250, 1977.
- [181] Lauren A Rivera. When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring. *American Sociological Review*, 82(6):1111–1138, 2017.
- [182] Richard N Landers. Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming*, 45(6):752–768, 2014.
- [183] Filip Lievens and Fiona Patterson. The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology*, 96(5):927, 2011.
- [184] Stephan J Motowidlo, Marvin D Dunnette, and Gary W Carter. An alternative selection procedure: The low-fidelity simulation. *Journal of applied psychology*, 75(6):640, 1990.
- [185] Janneke K Oostrom, Reinout E de Vries, and Mariska De Wit. Development and validation of a hexaco situational judgment test. *Human Performance*, 32(1):1–29, 2019.
- [186] Michael A McDaniel Cabrera and Nhung T Nguyen. Situational judgment tests: A review of practice and constructs assessed. *International journal of selection and assessment*, 9(1-2):103–113, 2001.
- [187] George C Thornton III and Deborah E Rupp. *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Psychology Press, 2006.
- [188] Herman Aguinias, Christine A Henle, and James C Beaty Jr. Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment*, 9(1-2):70–83, 2001.

- [189] International Task Force on Assessment Center Guidelines. Guidelines and ethical considerations for assessment center operations 1. *International Journal of Selection and Assessment*, 17(3):243–253, 2009.
- [190] Carrie A Blair, Brian J Hoffman, and Robert T Ladd. Assessment centers vs situational judgment tests: Longitudinal predictors of success. *Leadership & Organization Development Journal*, 2016.
- [191] Hamza Aouam, Bouchra Belmouffeq, and Aziza Mahil. Applying game-based approaches in personnel selection: a systematic literature review. *Management Dynamics in the Knowledge Economy*, 11(1):16–29, 2023.
- [192] Jim Lumsden, Elizabeth A Edwards, Natalia S Lawrence, David Coyle, Marcus R Munafò, et al. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR serious games*, 4(2):e5888, 2016.
- [193] Yigal Attali and Meirav Arieli-Attali. Gamification in assessment: Do points affect test performance? *Computers & Education*, 83:57–63, 2015.
- [194] Deborah L Whetzel and Michael A McDaniel. Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3):188–202, 2009.
- [195] James Clevenger, Gloria M Pereira, Darin Wiechmann, Neal Schmitt, and Veronica Schmidt Harvey. Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86(3):410, 2001.
- [196] Don C Zhang and Yi Wang. An empirical approach to identifying subject matter experts for the development of situational judgment tests. *Journal of Personnel Psychology*, 2021.
- [197] Fiona Patterson and Rachel Driver. Situational judgement tests (sjts). *Selection and recruitment in the healthcare professions: research, theory and practice*, pages 79–112, 2018.
- [198] Hartmut Koenitz. Towards a specific theory of interactive digital narrative. In *Interactive digital narrative*, pages 91–105. Routledge, 2015.
- [199] Janet H Murray. *Hamlet on the Holodeck, updated edition: The Future of Narrative in Cyberspace*. MIT press, 2017.

- [200] Chris Crawford. *Chris Crawford on interactive storytelling*. New Riders, 2012.
- [201] Nick Montfort. *Twisty Little Passages: an approach to interactive fiction*. Mit Press, 2005.
- [202] Brenda Kay Laurel. *TOWARD THE DESIGN OF A COMPUTER-BASED INTERACTIVE FANTASY SYSTEM (DRAMA, PLAYWRITING, POETICS, EXPERT SYSTEMS, THEORY)*. The Ohio State University, 1986.
- [203] Joseph Bates. Virtual reality, art, and entertainment. *Presence: Teleoperators & Virtual Environments*, 1(1):133–138, 1992.
- [204] Michael Mateas and Andrew Stern. Structuring content in the façade interactive drama architecture. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 1, pages 93–98, 2005.
- [205] David Thue, Vadim Bulitko, Marcia Spetch, and Eric Wasylishen. Interactive storytelling: A player modelling approach. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 3, pages 43–48, 2007.
- [206] Brian Magerko, John E Laird, Mazin Assanie, Alex Kerfoot, Devvan Stokes, et al. Ai characters and directors for interactive computer games. In *AAAI*, pages 877–883, 2004.
- [207] Boyang Li and Mark Riedl. An offline planning approach to game plotline adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 6, pages 45–50, 2010.
- [208] Ian Millington and John Funge. *Artificial intelligence for games*. CRC Press, 2009.
- [209] James Skorupski, Lakshmi Jayapalan, Sheena Marquez, and Michael Mateas. Wide ruled: A friendly interface to author-goal based story generation. In *Virtual Storytelling. Using Virtual Reality Technologies for Storytelling: 4th International Conference, ICVS 2007, Saint-Malo, France, December 5-7, 2007. Proceedings 4*, pages 26–37. Springer, 2007.
- [210] Mark Owen Riedl and Vadim Bulitko. Interactive narrative: An intelligent systems approach. *Ai Magazine*, 34(1):67–67, 2013.

- [211] Katherine Isbister, Kristina Höök, Michael Sharp, and Jarmo Laaksolahti. The sensual evaluation instrument: developing an affective evaluation tool. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1163–1172, 2006.
- [212] Tracy Fullerton. *Game design workshop: a playcentric approach to creating innovative games*. CRC press, 2014.
- [213] Clinton Keith. *Agile game development with Scrum*. Pearson Education, 2010.
- [214] Roy D Pea. User centered system design: new perspectives on human-computer interaction. *Journal educational computing research*, 3(1):129–134, 1987.
- [215] Andrew Brownsword. Reflecting on development processes in the video game industry. In *2009 31st International Conference on Software Engineering-Companion Volume*, pages 182–182. IEEE, 2009.
- [216] Jérôme Hernandez, Mathieu Muratet, Matthis Pierotti, and Thibault Carron. How to model a visual novel game to train and identify players’ soft skills? In *European Conference on Game Based Learning*, pages 322–330, 2021.
- [217] Allan Cameron. Modular narratives in contemporary cinema. In *Modular Narratives in Contemporary Cinema*, pages 1–19. Springer, 2008.
- [218] Emily Short. Beyond branching: quality-based, salience-based, and waypoint narrative structures. *Emily Short’s interactive storytelling*, 2016.
- [219] Wim J Linden, Wim J van der Linden, and Cees AW Glas. *Computerized adaptive testing: Theory and practice*. Springer, 2000.
- [220] Michael C Ashton and Kibeom Lee. The hexaco–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, 91(4): 340–345, 2009.
- [221] Jérôme Hernandez, Mathieu Muratet, and Thibault Carron. Responsive narrativity in gamified situational judgment tests: A dynamic system for enhanced soft skills development. In *European Conference on Technology Enhanced Learning*, pages 725–730. Springer, 2023.

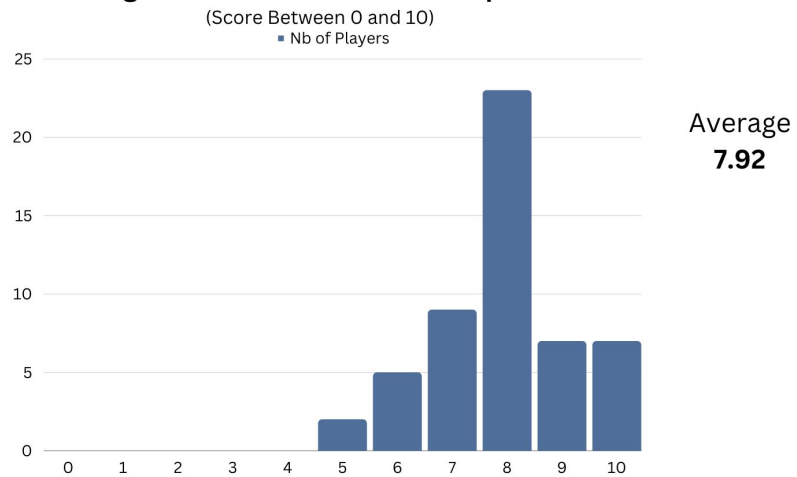
- [222] Qingxiong Derek Weng, Hui Yang, Filip Lievens, and Michael A McDaniel. Optimizing the validity of situational judgment tests: The importance of scoring methods. *Journal of Vocational Behavior*, 104:199–209, 2018.
- [223] Qingxiong Derek Weng, Hui Yang, Filip Lievens, and Michael A McDaniel. Optimizing the validity of situational judgment tests: The importance of scoring methods. *Journal of Vocational Behavior*, 104:199–209, 2018.
- [224] Mindy E Bergman, Fritz Drasgow, Michelle A Donovan, Jaime B Henning, and Suzanne E Juraska. Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, 14(3):223–235, 2006.
- [225] Filip Lievens, Helga Peeters, and Eveline Schollaert. Situational judgment tests: A review of recent research. *Personnel Review*, 37(4):426–441, 2008.
- [226] Frederick L Oswald, Neal Schmitt, Brian H Kim, Lauren J Ramsay, and Michael A Gillespie. Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of applied psychology*, 89(2):187, 2004.
- [227] Deborah L Whetzel, Michael A McDaniel, and Nhung T Nguyen. Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21(3):291–309, 2008.
- [228] Paul R Sackett and Steffanie L Wilk. Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, 49(11):929, 1994.
- [229] Keith Topping. Self and peer assessment in school and university: Reliability, validity and utility. *Optimising new modes of assessment: In search of qualities and standards*, pages 55–87, 2003.
- [230] Tomáš Kliegr, Štěpán Bahník, and Johannes Fürnkranz. A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artificial Intelligence*, page 103458, 2021.
- [231] Karolina Rab-Kettler and Bada Lehnervp. Recruitment in the times of machine learning. *Management Systems in Production Engineering*, 2019.

- [232] Jérôme Hernandez, Mathieu Muratet, Matthis Pierotti, and Thibault Caron. Enhancement of a gamified situational judgment test scoring system for behavioral assessment. In *2022 International Conference on Advanced Learning Technologies (ICALT)*, pages 374–378. IEEE, 2022.
- [233] Barbaros Bostan and Tim Marsh. Fundamentals of interactive storytelling. *AJIT-e*, 3(8):19, 2012.
- [234] Raul Paradedda, Maria José Ferreira, Raquel Oliveira, Carlos Martinho, and Ana Paiva. The role of assertiveness in a storytelling game with persuasive robotic non-player characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 453–465, 2019.
- [235] Joseph Bates et al. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.
- [236] Md Abdur Rahman, Asif Al Faisal, Tayeba Khanam, Mahfida Amjad, and Md Saeed Siddik. Personality detection from text using convolutional neural network. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pages 1–6. IEEE, 2019.
- [237] Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2):74–79, 2017.
- [238] Magalie Ochs, Nicolas Sabouret, and Vincent Corruble. Simulation of the dynamics of nonplayer characters’ emotions and social relations in games. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4):281–297, 2009.
- [239] John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.
- [240] PDB PTE LTD. Personality database™. retrieved from <https://www.personality-database.com>, 2022.
- [241] Brandon Cui and Calvin Qi. Survey analysis of machine learning methods for natural language processing for mbti personality type prediction, 2017.
- [242] Alam Sher Khan, Hussain Ahmad, Muhammad Zubair Asghar, Furqan Khan Saddozai, Areeba Arif, and Hassan Ali Khalid. Personality classification

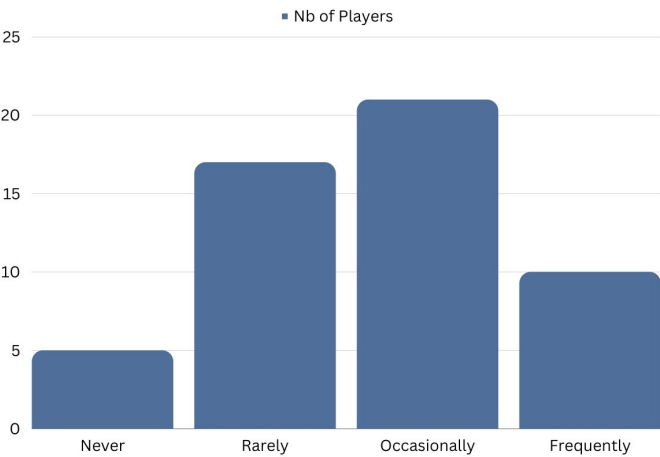
- from online text using machine learning approach. *International Journal of Advanced Computer Science and Applications*, 11(3):460–476, 2020.
- [243] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- [244] Ye Zhang and Byron Wallace. A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015.
- [245] David Talbot and Miles Osborne. Smoothed bloom filter language models: Tera-scale lms on the cheap. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 468–476, 2007.
- [246] Shristi Chaudhary, Ritu Singh, Syed Tausif Hasan, and Ms Inderpreet Kaur. A comparative study of different classifiers for myers-brigg personality prediction model. *Linguistic analysis*, page 21, 2013.
- [247] Jérôme Hernandez, Mathieu Muratet, Matthis Pierotti, and Thibault Caron. Can we detect non-playable characters’ personalities using machine and deep learning approaches? In *ECGBL 2022 16th European Conference on Game-Based Learning*. Academic Conferences and publishing limited, 2022.

Experimentation 1: Survey Questions

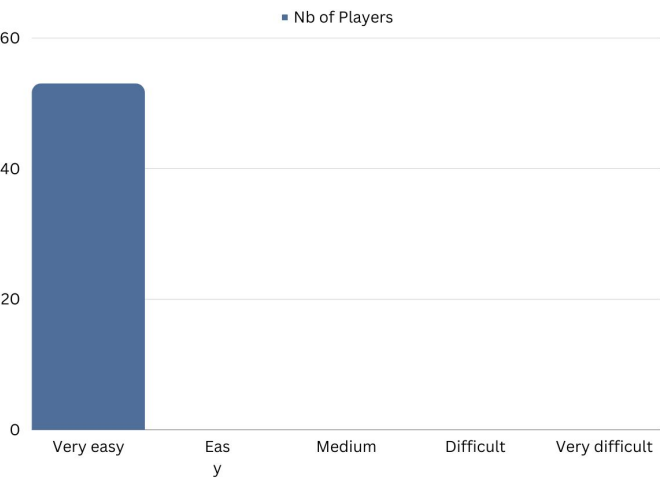
Q1: Did you enjoy this gamified experience focused on learning and raising awareness about leadership soft skills?



Q2: Have you ever experienced similar interpersonal situations to those in the game, either at work or during your studies?



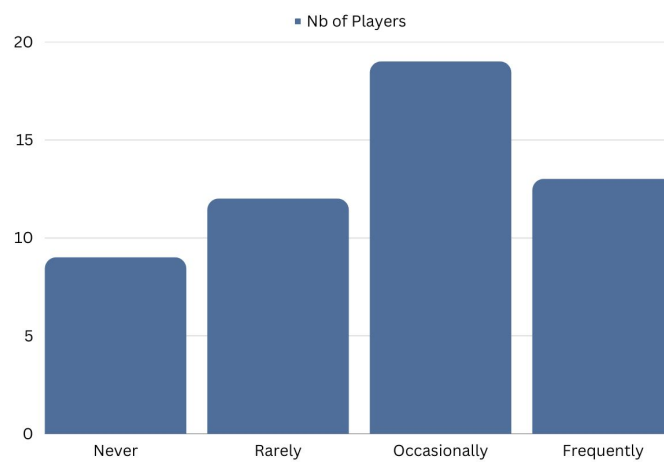
Q3: How did you find the game's user interface, controls, and gameplay?



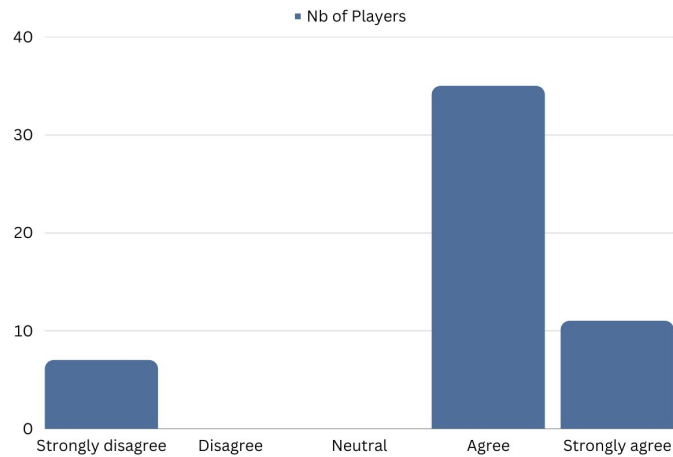
Q4: In a few lines, which aspects of the experience did you appreciate?

Q5: In a few lines, which aspects of the experience would you like to see improved and why?

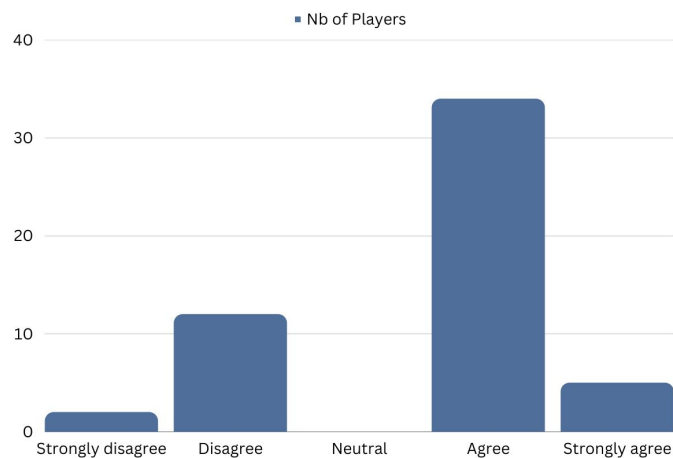
Q6: Do you play video games?



Q7: Do you think that this type of game can help improve performance at work or in the context of your studies when facing similar situations?



Q8: Do you believe that this type of game can influence your future behavior at work or in the context of your studies?



Q9: Are you a Woman, a Man or Other?